Data-Driven Fairness-Aware Vehicle Displacement for Large-Scale Electric Taxi Fleets

Guang Wang[1], Shuxin Zhong[1], Shuai Wang[2], Fei Miao[3], Zheng Dong[4], Desheng Zhang[1]
[1]Rutgers University, [2]Southeast University, [3]University of Connecticut, [4]Wayne State University
{guang.wang, sz517}@rutgers.edu, shuai.wang@seu.edu.cn, fei.miao@uconn.edu, dong@wayne.edu, dz220@cs.rutgers.edu

***Abstract*—We are witnessing a rapid taxi electrification process due to the ever-increasing concern about urban air quality and energy security. A key difference between conventional gas taxis and electric taxis is their energy replenishment mechanisms, i.e., refueling or charging, which is reflected in two aspects: (i) much longer charging processes vs. short refueling processes and (ii) time-varying electricity prices vs. time-invariant gasoline prices during a day. The complicated charging issues (e.g., long charging time and dynamic charging pricing) potentially reduce electric taxis' daily operation time and profits, and also cause overcrowded charging stations during some off-peak charging pricing periods. Motivated by a set of findings obtained from a data-driven investigation, in this paper, we design a fairness-aware vehicle displacement system called `FairMove` to improve the overall profit efficiency and profit fairness of electric taxi fleets by considering both the passenger travel demand and taxi charging demand. We first formulate the electric taxi displacement problem as multi-agent deep reinforcement learning, and then we propose a centralized multi-agent actor-critic approach to tackle this problem. More importantly, we implement and evaluate `FairMove` with real-world streaming data from the Chinese city Shenzhen, including GPS data and transaction data from more than 20,100 electric taxis, coupled with the data of 123 charging stations, which constitute, to our knowledge, the largest all-electric taxi network in the world. The extensive experimental results show that our fairness-aware `FairMove` effectively improves the profit efficiency and profit fairness of the Shenzhen electric taxi fleet by 25.2% and 54.7%, respectively.**

I. Introduction

Due to the ever-growing concern about air quality and energy security, more and more countries and cities have started their electric vehicle initiatives [1]. 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 [2]. As one of the most common mobility modes, taxis play an important role in people's daily life, and they are also in the front line of vehicle electrification given their long daily travel distances [3], [4]. For example, the number of electric taxis (e-taxi) in the Chinese city Shenzhen has increased from 840 in 2015 to over 20,130 in 2020. For the e-taxi fleet, one of the most important tasks for a fleet management team to improve fleet efficiency and resultant fleet profit is **Vehicle Displacement**, i.e., to make recommendations to individual vacant e-taxis (i.e., without passengers) to proactively go from one area to another area to balance two relationships: the future passenger demand and supply; and the future e-taxi charging demand and supply.

Admittedly, a large number of works have been done to improve the efficiency of taxi fleets [5]–[7]. However, a majority of these works focused on *conventional gas taxis*. For example, [6], [7] in the SIGSPATIAL Cup 2019 tried to optimize the efficiency of taxi drivers searching for customers, e.g., minimize the average searching time. Different from the refueling process of gas taxis, which usually takes about 3-5 minutes, the charging process of e-taxis typically lasts for half an hour to two hours even with fast chargers [8]. In addition, the electricity price is varying in different hours of the day, while the gasoline price is usually constant during the day. These long charging time and dynamic charging pricing lead to very different e-taxi drivers' behaviors (where or when to charge) and incentive (whether to follow a recommendation). Even though the energy level can be naively considered as a constraint of existing solutions to address the charging problem (e.g., if the energy level of an e-taxi is lower than a threshold, the e-taxi is set to be offline and removed from the system, so it is similar to passenger searching), the key challenge is to decide which charging station the e-taxi should go. The charging scheduling decisions are related to many factors like real-time traffic conditions, the status of charging stations and charging prices, which should be considered for reducing the charging idle time and charging costs (e.g., avoid overcrowded charging stations). In [6], [7], they considered the taxis are purely competitive, which potentially causes profit inequality between taxis and system unsustainability. However, overall fairness should be achieved in a cooperative fashion. Hence, the existing solutions for vehicle relocation or passenger searching [6], [7] are not suitable to the fairness-aware vehicle displacement problem well. Recently, some works have been done to understand and improve the *charging efficiency of e-taxis* [1], [8], [9], but almost all of them mainly focus on optimizing charging idle time reduction instead of optimizing drivers' profits due to a lack of detailed transaction data from taxi fleets. The underlying assumption by the existing works is that the optimizing charging idle time will prolong the e-taxis' operation time. However, we found that *charging idle time reduction does not necessarily indicate the prolonged time for serving passengers to maximize profit because some drivers may need to spend more time to seek passengers after charging in some regions with shorter charging waiting time but lower passenger demand* (as shown in Section II-C).

In this paper, by working with an e-taxi agency in China, we utilize its detailed proprietary transaction data along with

TABLE I
AN EXAMPLE OF SOME KEY FIELDS OF THE DATASETS.

E-Taxi GPS	E-Taxi Transaction	Charging Station	Urban Partition	Charging Pricing
vehicleID: YBD00000	vehicleID: YBD00000	stationID: 53	regionID: 1	low rate (CNY/kWh): 0.9
longitude: 114.08663	pickupTime: 2019-12-11 15:27:24	stationName: LGS0012	longitude1: 114.31559	flat rate (CNY/kWh): 1.2
latitude: 22.63244	dropoffTime: 2019-12-11 16:11:32	longitude: 114.07051	latitude1: 22.78559	high rate (CNY/kWh): 1.6
time: 2019-12-11 06:58:55	cost (CNY): 54	latitude: 22.65589	longitude2: 114.31123	–
speed (km/h): 27	distance (km): 16.2	# of charging points: 52	latitude2: 22.78220	–

GPS data to improve the e-taxi fleet's profit efficiency by designing a new vehicle displacement system called `FairMove`, which balances two relationships: passenger demand vs. e-taxi supply, and vehicle charging demand vs. charging station supply. For the e-taxi agency, its goal for vehicle displacement is to optimize the overall profit efficiency of the e-taxi fleet; whereas for the drivers, their incentive to following vehicle displacement is to enable profit fairness among them. As a result, the key objective of our e-taxi displacement is fairness-aware fleet-wide profit efficiency optimization.

However, the e-taxi displacement with this objective is challenging due to possible conflicting relationships (e.g., balancing future passenger demand and supply vs. balancing future e-taxi charging demand and supply), and many confounding factors (e.g., individual drivers charging behaviors like spatiotemporal charging preference, time-variant charging pricing, and individual-level fairness). To address these challenges, in this paper, we propose a deep reinforcement learning (DRL)-based approach (i.e., Centralized Multi-Agent Actor-Critic (CMA2C for short) to learn the sophisticated e-taxi displacement policy. `CMA2C` has three key advantages for the e-taxi displacement: (i) displacement decisions for e-taxis are sequential and highly repetitive, thus generating an abundance of training data for training `CMA2C` algorithms (e.g., displacement actions and rewards); (ii) `CMA2C` can approximate the Q-value function (e.g., Deep Q-Networks (DQN)) [10] and learn complex decision-making policies by deep neural networks, which has the capability to deal with the e-taxi displacement problem with huge state space; (iii) `CMA2C` maximizes the long-term reward of a sequence of decisions for profit efficiency and profit fairness improvement.

In particular, the key contributions of this paper include:

- We conduct an extensive data-driven analysis based on real-world multi-source data, from which we found some novel insights: (i) Charging time reduction does not necessarily indicate the prolonged time for serving passengers since some drivers may need to spend more time to seek passengers after charging in some regions with low passenger travel demand. (ii) The potential profits for serving passengers after charging in different stations may also be different, which is highly dynamic in both spatial and temporal dimensions. (iii) Prolonged charging time of e-taxis compared to the refueling processes of gas taxis causes some real-world issues, e.g., intensive charging peaks and long charging wait time in some time slots induced by the time-varying charging pricing.
- Based on the data-driven insights, we design a new fairness-aware displacement system called `FairMove` to improve the overall profit efficiency and profit fairness for e-taxi fleets by a CMA2C approach. `FairMove` considers not only the operation behaviors of drivers and demand & supply but also the complicated charging processes (e.g., time-varying charging pricing, and intensive charging peaks). In addition, the time for seeking a passenger after charging and trip length are also considered for a more accurate revenue estimation. Finally, both the operating revenues and charging costs are fed to the `FairMove` system to make fair-profit oriented decisions, which has the potential to make the system more sustainable.
- More importantly, we implement and extensively evaluate our `FairMove` based on multi-source data from the Chinese city Shenzhen, including GPS records and transaction records from 20,130 e-taxis. The experimental results show our `FairMove` effectively increases the profit efficiency of the e-taxi fleet by 25.2%, improves the profit fairness of e-taxi drivers by 54.7%, and reduces the cruise time and idle time by 32.1% and 43.3% on average at the same time.

II. DATA AND MOTIVATION

A. Data Description

All e-taxis in Shenzhen are the same vehicle model, i.e., BYD e6, whose battery capacity and maximum traveling distance are 80 kWh and 400 km, respectively [11]. There are five datasets used in our paper, i.e., the e-taxi GPS data, the e-taxi transaction data, the charging station data, urban partition data, and the time-variant electricity rates data. The detailed information of the five datasets is shown as follows.

(i) GPS data includes vehicle IDs, real-time coordinates (i.e., longitudes and latitudes), time stamps, directions, speeds, and passenger indicator.

(ii) Transaction fare data includes vehicle IDs, the pickup and drop-off times, the pickup and drop-off coordinates (i.e., longitudes and latitudes), operating distances, cruising distances, and fares.

(iii) Charging station data includes station IDs, station names, coordinates (i.e., longitudes and latitudes), and the number of fast charging points in each station. There are 123 charging stations deployed in Shenzhen for e-taxis only in December 2019.

(iv) Urban Partition Data describes the urban partition for the population census of the Chinese city Shenzhen, which is provided by the Shenzhen government. There are 491 regions,

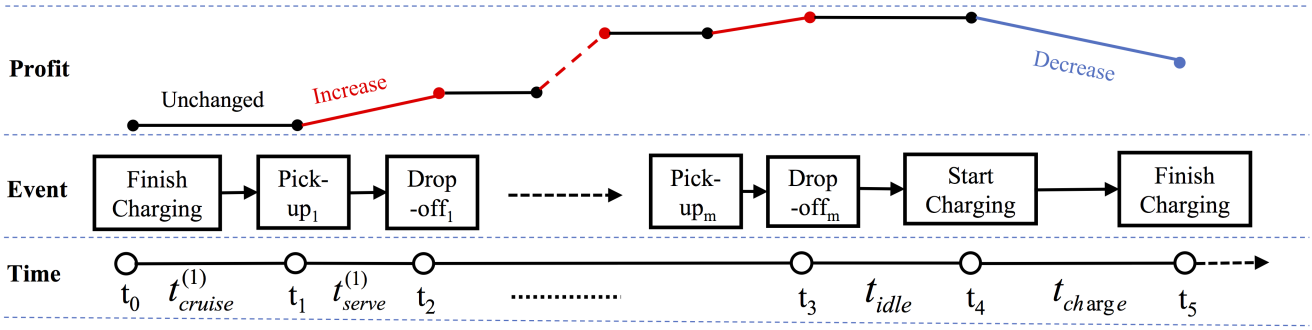

Fig. 1. Mobility decomposition of e-taxis.

and each region has a region ID and longitudes & latitudes of its boundary.

(v) Charging Pricing Data. Many cities have time-variant charging pricing (similar to the time-variant electricity pricing), which breaks up 24 hours of a day into several intervals and charges a different price for each interval [12]. 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 charging rates are 0.9, 1.2, and 1.6 CNY/kWh, respectively. The time-variant charging pricing in Shenzhen is shown as Fig. 2. An example of some key fields of the other four datasets is shown in Table I.

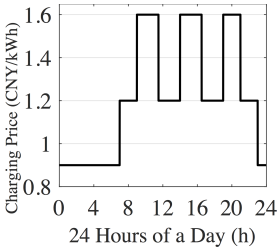

Fig. 2. Charging Prices of E-Taxis in Shenzhen.

B. Mobility Decomposition of E-Taxis

We depict the mobility of e-taxis from three dimensions by charging events, as shown in Fig. 1, where t_0 to t_5 represents the activities of an e-taxi during two consecutive charging events.

(i) At the time t_0, an e-taxi finishes a charging event, and then it will cruise to find passengers to serve. At t_1, the e-taxi picks the first passenger up and drops the passenger off at time t_2. We define the time for seeking a passenger as the **cruise time**, and the time for serving a passenger (onboard) as the **service time**. Specifically, we define the time duration $t_1 - t_0$ as the *first cruise time* $t^{(1)}_{cruise}$, and the time duration $t_2 - t_1$ as the *first service time* $t^{(1)}_{serve}$. During the *cruise time*, the e-taxi neither has passengers on board nor charges, so the profit remains unchanged. During the *service time*, the e-taxi's profit will increase with passengers on board. The profit is typically a function of time and distance.

(ii) After serving the first passenger, the e-taxi will continue to cruise and serve the $2^{nd}, 3^{rd}, \ldots, m^{th}$ passenger, and the e-taxi's profit keeps increasing during this period. After dropping the m^{th} passenger off, the energy level of this e-taxi decreases to a threshold, so it will start to seek a charging station to charge at time t_3. We define time duration $t_3 - t_0$ as the **operation time** T_{op}, which equals to $T_{cruise} + T_{serve}$, where $T_{cruise} = \sum_{i=1}^{n} t^{(i)}_{cruise}$ and $T_{serve} = \sum_{i=1}^{n} t^{(i)}_{serve}$. The profit of the e-taxi keeps increasing during T_{op} for continuously serving passengers.

(iii) Due to some real-world issues (e.g., inadequate charging resources and intensive charging peaks), the e-taxi may need to wait for a while to get an available charging point. Then at time t_4, there is an available charging point, so the driver will plug in the charger and charge the taxi. We define time duration $t_4 - t_3$ as the **idle time** T_{idle} since the e-taxi neither operates nor charges. The profit of the e-taxi remains unchanged during the *idle time*.

(iv) After plugging in a charger, the e-taxi will start to charge, and it finishes the charging event at time t_5. We define the time duration with a charger plugin $t_5 - t_4$ as the **charge time** T_{charge}. During this time period, the profit of the e-taxi will decrease due to energy replenishment.

(v) We define the time duration between two sequential charging events $t_5 - t_0$ as a **working cycle** T_{cycle} of an e-taxi, which equals to $T_{op} + T_{idle} + T_{charge}$. Hence, during a long time period (e.g., one week), there will be a set of *working cycles* for each e-taxi. In this paper, we focus on the long-term (e.g., weekly) profit fairness of e-taxis instead of the short-term profit, which also has the potential to achieve a higher overall profit efficiency for e-taxi fleets.

C. Motivation By Data-Driven Findings

Based on our multi-source real-world data and the above definitions, we conduct in-depth data-driven analysis using one-month e-taxi data to show the uniqueness and motivation of our e-taxi displacement design. In particular, we provide the following findings:

(i) We found the charging time of e-taxis is very long compared to the refueling time of gas taxis. Specifically, the charging time of 73.5% of charging events lasts for 45 minutes to two hours, as shown in Fig. 3, which is much longer than the refueling processes of conventional gas taxis (usually 3-5 minutes). Hence, the e-taxi displacement would be different from the conventional gas taxi dispatching considering the complicated charging issues, which results in that the existing dispatching strategies for gas taxis may not be suitable for the displacement of e-taxi fleets.

(ii) Due to the operation patterns and time-varying charging pricing (caused by time-varying electricity pricing), there are intensive charging peaks during some low charging pricing durations (e.g., 2:00-6:00, 12:00-14:00, and 17:00-18:00 as

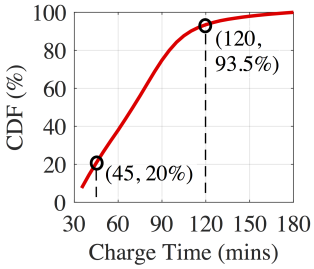

Fig. 3. Charging duration.

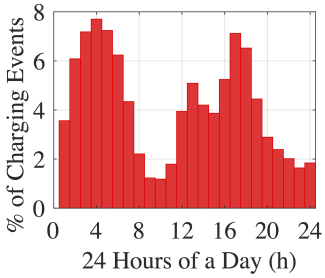

Fig. 4. Charging temporal pattern.

shown in Fig. 4), which causes that some charging stations to be overcrowded and potentially prolong the *idle time* of e-taxis given the long charge time. Hence, compared to the time-invariant gasoline prices, the time-varying charging prices also make a difference between e-taxis and gas taxis, which potentially makes the e-taxi displacement problem more complicated compared to the gas taxi dispatching.

(iii) *Idle time* reduction does not necessarily indicate the prolonged time for serving passengers since some e-taxis may need to spend more time seeking passengers after charging in some regions with low passenger travel demand. As shown in Fig. 5, we found 40% of e-taxis can find their first passengers after charging in 10 minutes, but there are still 10% of e-taxis need to cruise over an hour to find their first passenger after charging. In addition, the *first cruise time* $t^{(1)}_{cruise}$ is also different when charging in different stations. Fig. 6 shows the *first cruise time* of e-taxis after charging in three different charging stations. The three charging stations are located in different areas of the city, and there are a different number of charging points in each station. We found that the *first cruise time* of the e-taxis has large differences after charging in different stations. Hence, the charging station selection not only impacts the *idle time* but also has influences on the *first cruise time* $t^{(1)}_{cruise}$. However, this finding has not been revealed and considered by existing works.

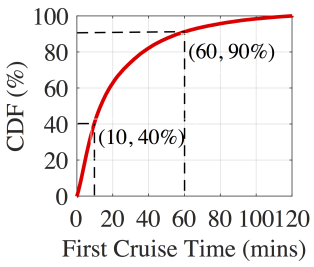

Fig. 5. First cruise time distribution.

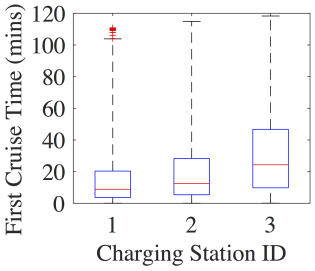

Fig. 6. First cruise time of three charging stations.

(iv) The potential revenue for serving passengers after charging may also be different at different time slots and stations, which is highly dynamic in both spatial and temporal dimensions. Fig. 7 shows a visualization of the average per-trip revenue in different regions during late night (00:00-01:00), morning rush hour (08:00-09:00), and evening rush hour (18:00-19:00). The dark red means higher average per-trip revenue (i.e., more long trips) in these regions, and light yellow means lower average per-trip revenue in these regions. We found the average per-trip revenue has a large gap between different regions across the city, ranging from several CNY to over 100 CNY. For example, the per-trip revenue in the airport region is always high, but it is very low in some suburban areas. In addition, we found the average trip length in a region may change during the day. We also quantify the average per-trip revenue in the 491 regions, which can be seen from the right upper corner of Fig. 7. We found that there are more regions with low prices per trip during the late-night, but more regions with high prices per trip during rush hours.

Certainly, the passenger travel demand and supply in different regions are also different, so the probability to pick up a passenger is also different, which is usually considered by existing works. However, existing works rarely consider the revenue from a trip, which will also directly impact taxis' revenue. Hence, in this paper, we consider not only the demand and supply of e-taxis but also the potential revenue for serving passengers for displacement, which lays a foundation for e-taxi' revenue fairness.

(v) Inequal profit efficiency of the e-taxi fleet, which could be potentially improved by a centralized displacement system. As shown in Fig. 8, we found 20% of e-taxis' hourly profit efficiency is lower than 36, and there are also 20% of e-taxis' hourly profit efficiency is higher than 51, which means there is a huge profit gap between e-taxis, resulting in the profit of high-efficient drivers will be 42% higher than the low-efficient drivers. With people pay more attention to fairness and equity, such a large profit gap potentially hurts some drivers' daily life and makes them unsatisfactory. Hence, it is necessary to have a centralized fairness-aware displacement system to improve e-taxis' profit fairness in the fleet without damaging the overall profit efficiency of the fleet.

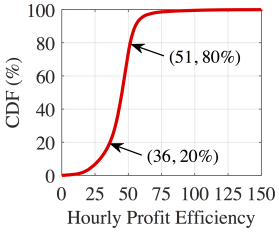

Fig. 8. Hourly profit efficiency.

In summary, based on our data-driven observations, we found the e-taxi displacement problem is also different from the existing charging scheduling/recommendation since (i) the *idle time* reduction does not necessarily indicate more time for serving passengers. (ii) The *first cruise time* of the e-taxis has large differences after charging in different stations. (iii) Not only the probability of picking up passengers impacts e-taxis' revenue but also the trip length has a huge impact on it, which we could also consider improving the profit fairness for the e-taxi fleet. (iv) It is necessary to design a fairness-aware displacement system for e-taxi fleets to improve their profit efficiency and fairness, but it may not be achieved by existing solutions for conventional gas taxis.

III. FairMove Displacement System Design

In this section, (i) we first show the key idea of the FairMove design, (ii) followed by the problem statement. (iii) We then formulate the e-taxi displacement problem as *collaborative multi-agent Markov decision process*, which

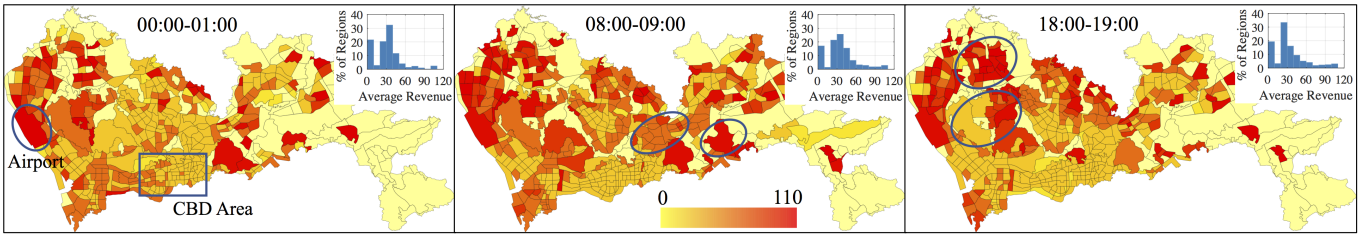

Fig. 7. Average per-trip revenue (CNY) in different regions during different hours of a day.

consists of multiple agents cooperating to achieve one goal. (iv) Then we propose a multi-agent DRL approach called Centralized Multi-Agent Actor-Critic (i.e., CMA2C) to tackle this problem, which effectively utilizes large-scale historical data to train the DRL agents.

A. Key Idea of `FairMove`

The key idea of our `FairMove` is that we formulate the taxi displacement problem as a large-scale sequential decision-making problem since the displacement decisions for e-taxis are sequential and highly repetitive, where each decision corresponds to scheduling an available (vacant) e-taxi to a region or a charging station. There are multiple, possibly conflicting objectives in our displacement system, e.g., improving the profit efficiency of the e-taxi fleet and reducing the profit unfairness between all e-taxis. In this work, we balance the profit efficiency and profit fairness by a weighted parameter to achieve the optimal displacement policy.

However, it is challenging to design an effective displacement strategy for e-taxi fleets that can adapt to an environment involving dynamic demand & supply and complicated charging behaviors as shown in the above data-driven investigation. One major issue is that changes in a displacement decision will impact future demand & supply, and it is challenging for supervised learning approaches to capture and model these real-time changes. Inspired by successful applications in intellectual challenging decision-making problems (e.g., the game of Go [13], worker scheduling [14]), in this paper, we try to target the e-taxi displacement problem by deep reinforcement learning (DRL) based methods, which combine the advantages of Deep Neural Networks (DNNs) and Reinforcement Learning (RL) and has the capability of handling high-dimension data and highly dynamic environment features.

B. Problem Statement

Definition 1. (Profit Efficiency) Profit efficiency denotes the per unit time profit earned by an e-taxi during its on-duty time in a period Γ (e.g., a week). The on-duty time of an e-taxi includes a set of *working cycles*. Each *working cycle* consists of three components, i.e., operation time T_{op}, idle time T_{idle}, and charging time T_{charge}. The calculation method of the Profit efficiency of each e-taxi can be represented as Equation 1.

$$PE = \frac{\text{Revenue} - \text{Costs}}{\sum_{k=1}^{z} T_{cycle}^{(k)}} = \frac{\sum_{i=1}^{m} R_{trip}^{(i)} - \sum_{j=1}^{n} C_{charge}^{(j)}}{\sum_{k=1}^{z} \left(T_{op}^{(k)} + T_{idle}^{(k)} + T_{charge}^{(k)}\right)} \quad (1)$$

where PE denotes the *Profit Efficiency* of an e-taxi in a period Γ. *Revenue* and *Costs* denote the total revenue earned from serving passengers and the operation costs during Γ by the e-taxi, respectively. m, n, and z denotes the number of trips served by the e-taxi, the number of charging events of the e-taxi, and the number of working cycles of the e-taxi during Γ. $R_{trip}^{(i)}$ is the revenue for serving ith trip, $C_{charge}^{(j)}$ is the charging cost for jth charging event. $T_{op}^{(k)}$, $T_{idle}^{(k)}$, $T_{charge}^{(k)}$ are the operation time T_{op}, idle time T_{idle}, and charge time T_{charge} of kth *working cycle*, respectively.

In this paper, since we define a *working cycle* as the time between two charging events, so $z = n$ in Equation 1, and $T_{op}^{(k)}$ is equivalent to $T_{cruise}^{(k)} + T_{serve}^{(k)}$. In addition, the charging costs is a function of the time-varying charging pricing and the charge time, so we describe the charge time of jth charging event $T_{charge}^{(j)}$ as a three-dimensional vector $T_{charge}^{(j)} = \left[T_p^{(j)}, T_f^{(j)}, T_o^{(j)}\right]$, where $T_p^{(j)}$, $T_f^{(j)}$, and $T_o^{(j)}$ denote the time in peak, flat, and off-peak charging pricing hours of the jth charging event, respectively. Similarly, we also describe the time-varying charging pricing as a three-dimensional vector $\lambda = [\lambda_p, \lambda_f, \lambda_o]$, where λ_p, λ_f, λ_o denote the charging prices during peak, flat, and off-peak hours, respectively (as shown in Fig. 2). Hence, we convert Equation 1 into Equation 2 to calculate the *profit efficiency* of an e-taxi.

$$PE = \frac{\sum_{i=1}^{m} R_{trip}^{(i)} - \sum_{j=1}^{n} \left(\lambda \cdot T_{charge}^{(j)}\right)}{\sum_{j=1}^{n} \left(T_{cruise}^{(j)} + T_{serve}^{(j)} + T_{idle}^{(j)} + T_{charge}^{(j)}\right)} \quad (2)$$

Definition 2. (Profit Fairness) It is typically challenging to define the fairness as different people may have different perceptions of fairness [15]. In addition, the fairness definition would also be different in different scenarios [16]. Hence, to better understand e-taxi drivers' perceptions to fairness and define the profit fairness of e-taxis properly, our team has conducted a set of interviews with Shenzhen e-taxi drivers and asked them related questions. We found almost all e-taxi drivers thought it is fair when their profits are proportional to their working time. Motivated by this, in this paper, we define the *Profit Fairness PF* of an e-taxi fleet as the variance of *profit efficiency* of all e-taxis in the fleet, which is denoted as Equation 3, so smaller PF means fairer for the e-taxi fleet.

$$PF = \frac{1}{N} \sum_{k=1}^{N} \left(PE^{(k)} - \overline{PE}\right)^2 \quad (3)$$

where N is the number of e-taxis in the e-taxi fleet. $PE^{(k)}$ is the *profit efficiency* of the kth e-taxi. $\overline{PE}$ is the average *profit efficiency* of all e-taxis in the fleet.

In this paper, we tackle the displacement problem for a large-scale available (i.e., vacant) e-taxi fleet under centralized management, considering both serving passengers and charging. The objectives of our displacement are three-fold: (1) Improving the overall *profit efficiency* PE of all e-taxis in the fleet during a period of Γ. (2) Enhancing the *profit fairness* of the e-taxi fleet PF over Γ. (3) Tradeoff between the *profit efficiency* and *profit fairness*.

For the spatial partition, we utilize the urban partition data described in Section II to represent the map, which splits the Shenzhen city into 491 regions. Our partition is similar to the grid-based methods (e.g., square-grid [1] and hexagonal-grid [17]), but our partition is more practical as it considers the geological structure of the city (e.g., a mountain or a lake will be partitioned in a single region). For the temporal partition, we split the duration of a day into T time slots. At each time slot, there are a different number of passenger demands sporadically appear in each region, and those passengers will be served by the available e-taxis in the same region. The role of the displacement system is to decide which region or charging station each vacant e-taxi should go in each time slot to maximize the future long-term *profit efficiency* and *profit fairness* of the e-taxi fleet.

C. Problem Formulation

Formally, we model the e-taxi displacement problem as a *multi-agent Markov decision process* $\mathcal{G}$ for N agents, which is defined by a five-tuple $\mathcal{G} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \beta)$, where $\mathcal{S}$ is the set of states; $\mathcal{A}$ is the joint action space; $\mathcal{P}$ is transition probability functions; $\mathcal{R}$ is the reward function; and β is a discount factor.

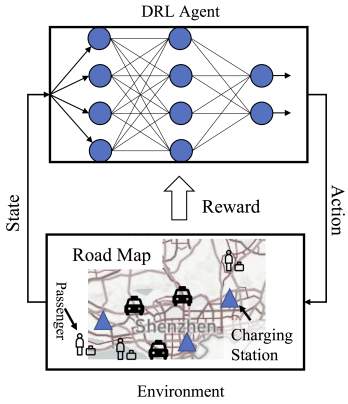

Fig. 9. A Markov decision process.

Markov decision process is typically used to model sequential decision-making problems. In a Markov decision process, an agent behaves in an environment to observe the state from the environment and makes action decisions for the environment to execute. Then, the environment feeds a reward for the decisions back to the agent and turns to the next state. The agent evolves along with this interaction with the environment. An example of the process can be seen in Fig. 9. The key of the Markov decision process is to produce an optimal policy that governs the decision-making at each step based on estimating the state-action value function of the agent. This function tells us how good a decision made at a particular location and time under given environment contexts with respect to the long-term objectives.

The detailed definitions of the *Markov decision process* $\mathcal{G}$ in our `FairMove` displacement system are shown as below.

Agent Set: We consider each available (i.e., vacant) e-taxi as an agent, and e-taxis within the same spatial-temporal partition are homogeneous, i.e., e-taxis in the same region or charging station during the same time slot are considered as homogeneous agents (where agents have the same states), and the number of available agents N_t is changing over time.

State $\mathcal{S}$: The state of an e-taxi k $s_t(k) \in \mathcal{S}$ consists of a two-dimensional vector indicating its specific spatiotemporal status from both the local view and global view. We discrete one day into a set of T time slots. And we divide the city into a set of R regions and C charging stations, (i.e., $R \cup C =$ the whole city; $R \cap C = \emptyset$). We define a local-view state of an e-taxi, $s_{t,lo} = [t, l] \in \mathcal{S}_{lo}$, where $t \in T$ is the time index (i.e., which time slot), and $l \in R \cup C$ is the location index (i.e., which region or charging station) where the e-taxi is in. In this case, the finite local state space $\mathcal{S}_{lo}$ is a Cartesian product of the set of time slots and the set of regions + charging stations, i.e., $\mathcal{S}_{lo} = T \times (R \cup C)$ and the number of states is $|\mathcal{S}_{lo}| = |T| \times |(R \cup C)|$. The e-taxis in the same partition (region or charging station) in a time slot have the same state. We also define a global-view state $s_{t,go}$, which is shared by all available e-taxis in the time slot t. The global-view state includes three different spatiotemporal features: (i) the number of available e-taxis in each region; (ii) the number of unoccupied charging points in each charging station; and (iii) the expected number of passengers in each region at the next time slot, which is predicted with historical and real-time data. The global-view state $s_{t,go}$ will update in each time slot. Finally, the state of each available e-taxi k during the time slot t can be represented as $s_t(k) = [s_{t,lo}(k), s_{t,go}(k)] \in \mathcal{S}(k)$. The joint state of all available e-taxis in the time slot t can be denoted as $\mathbf{s}_t \in \mathcal{S} = \mathcal{S}(1) \times \mathcal{S}(2) \times \ldots \mathcal{S}(N_t)$.

Action $\mathcal{A}$: The action space of an e-taxi k, $\mathcal{A}(k)$ specifies where it is able to arrive at the next time slot. There are three types of actions in our e-taxi displacement setting. (i) The first type of action is staying in the current region. (ii) The second type of action is displacing the e-taxi to another adjacent region in the direction of the potential nearest passenger. (iii) The third type of action is charging in a charging station. For the second type of action, each e-taxi can go to its adjacent regions and e-taxis in different regions have a various number of neighbor regions, so they have a different number of actions. The e-taxis in the same region have the same action space. For the third type of action, we consider the nearest five charging stations for each e-taxi to reduce the action space. The charging action is decided by the energy level of each e-taxi, which is estimated by the initial energy level and energy consumed for operating [16]. The energy consumption rate would be different if the e-taxis are made of various models with different characteristics, but the method is still applicable if we use separate energy consumption calculation formulas for them. At each time slot t, each available e-taxi takes an action $a_t(k) \in \mathcal{A}(k)$, forming the joint action $\mathbf{a}_t \in \mathcal{A} = \mathcal{A}(1) \times \mathcal{A}(2) \cdots \times \mathcal{A}(N_t)$, which induces a transition

in the environment according to the state transition function $\mathcal{P}\left(s_{t+1}|s_t, \mathbf{a}_t\right) : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$.

Reward $\mathcal{R}$: Reward reflects the immediate sense of the action in a specific state, but it is not equivalent to the goal. However, reward usually determines the optimization goal of the displacement system, which usually utilizes rewards to guide the learning process. A typical measurement is to estimate the difference of the accumulated reward between with and without an action. We define three types of immediate rewards in our e-taxi displacement scenario, i.e., positive rewards for serving passengers, 0 reward for cruising, and negative rewards for charging. Note that both the positive rewards and negative rewards are nondeterministic as the positive rewards are mainly decided by the trip length, and the negative rewards are decided by the charging time t_{charge} and time-variant charging prices. We consider the e-taxis will always serve the nearest passengers, and the passengers in a region will always be served by the vacant and available e-taxis. We define the reward as 0 when the e-taxi is cruising since there is no direct transaction. When the energy status of an e-taxi decreases to a certain threshold η (e.g., 20%), the e-taxi should go to charge. Even though the immediate reward for charging is negative, e-taxis cannot operate and serve passengers without energy, so running out of battery will cause no reward in the future. Hence, the charging action will also benefit the long-term positive reward for the e-taxi, which means the impact of charging can be spread to its future states. Considering both the profit efficiency and fairness, the final reward of the e-taxi k can be represented by Equation 4. $PE(k,t)$ is the profit efficiency of e-taxi k in the time slot t (i.e., in regard to state $s_t(k)$ and action $a_t(k)$), $PF(t)$ is the profit fairness of all active e-taxis in the time slot t. Since the PF Equation 3 indicates the unfairness of the system, so we have the minus here to maximize the profit fairness.

$$r(s_t(k), a_t(k)) = \alpha \cdot PE(k,t) + (1-\alpha)(-PF(t)) \tag{4}$$

To balance the profit efficiency and profit fairness, a real-valued parameter $\alpha \in [0,1]$ is leveraged to control how much we emphasize the profit efficiency and profit fairness of e-taxis in the fleet. As a boundary case, with $\alpha = 1$, we only explicitly maximize the profit efficiency for the fleet, while ignoring the level of unfairness among all e-taxis. With $\alpha = 0$, we only explicitly maximize the profit fairness for all e-taxis in the fleet, while ignoring the overall profit efficiency. Therefore, the Equation 4 can be converted to Equation 5.

$$r(k,t) = \alpha \cdot \frac{\sum_{i=1}^{m} R_{trip}^{(i)}(k,t) - \sum_{j=1}^{n}\left(\lambda \cdot T_{charge}^{(j)}(k,t)\right)}{\sum_{j=1}^{n}\left(T_{cruise}^{(j)}(k,t) + T_{serve}^{(j)}(k,t) + T_{idle}^{(j)}(k,t) + T_{charge}^{(j)}(k,t)\right)} + (1-\alpha) \cdot \left(-\frac{1}{N_t}\sum_{h=1}^{N_t}\left(PE(h,t) - \overline{PE(t)}\right)^2\right) \tag{5}$$

where we use $r(k,t)$ instead of $r(s_t(k), a_t(k))$ for short, shown as below. Our reward function considers both self profit efficiency and the fairness, so every e-taxi is not only maximizing its own profit when they learn the policy but also cooperating with each other to maximize the profit fairness, when every e-taxi is trying to maximize their expected discounted future rewards $\mathbb{E}\left[\sum_{i=0}^{\infty} \beta^i r(k, t+i)\right]$.

State transition function $\mathcal{P}$ is defined as a mapping $\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0,1)$. $p(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$ denotes the probability of transition to $\mathbf{s}_{t+1}$ given a joint action $\mathbf{a}_t$ in the current state $\mathbf{s}_t$. Notice that although the action is deterministic, the number of available e-taxis and passengers are different at different regions during each time slot.

Discount factor β essentially determines how much the reinforcement learning agents care about rewards in the distant future relative to those in the immediate future. The value of β is typically selected from $[0,1)$, so the final expected reward in the infinite horizon will be convergent and bounded to a finite number. If $\beta = 0$, the agent will be completely myopic and only learn about actions that produce an immediate reward without considering the future reward.

D. Centralized Multi-Agent Actor-Critic

In this section, we show how we solve the above formulated multi-agent problem with deep reinforcement learning-based method. There are typically two types of solutions: decentralized methods and centralized methods [18]. However, the decentralized methods are not well adapted to our problem for the following two reasons. (i) The number of available e-taxis keeps changing in different time slots, so it is challenging to identify the number of neural networks to be constructed. (ii) The travel demand is highly dynamic and the number of agents is extremely large, resulting in modeling each e-taxi with a separate neural network that has high computational costs. Therefore, in this paper, we resort to centralized methods, which utilize a single neural network to model the behaviors that are shared by all e-taxis, in other words, the parameters of the centralized neural network are shared by all e-taxis.

In this paper, we propose a centralized multi-agent actor-critic (CMA2C) algorithm to solve the above-defined multi-agent problem for large-scale e-taxi displacement, which is a multi-agent policy gradient algorithm that iterates its policy to adapt to the dynamically evolving action space. The basic idea of the CMA2C is that there are two networks, a policy network (i.e., Critic, which is utilized to output policy) and a value network (i.e., Actor, which is leveraged to evaluate the performance of the policy network). There are two key tasks for training the CMA2C, i.e., (i) learning the parameters of the policy network θ_p and (ii) learning the parameters of the value network θ_v. Both the Critic and Actor functions are parameterized with deep neural networks, and the parameters of the critic θ_p and actor θ_v are updated iteratively.

The centralized value function is shared by all active e-taxis with an expected update via observing the global state information, which means available e-taxis cooperatively take actions for a fairness-aware optimal strategy. The centralized state-value function is learned by minimizing the following loss function $\mathcal{L}\left(\theta_v\right)$ as shown in Equation 6, which is derived from the Bellman equation.

$$\mathcal{L}(\theta_v) = \left(\left(V_{\theta_v}(s_t(k)) \right) - V_{tg}(s_{t+1}(k); \theta_v^{'}, \pi) \right)^2 \quad (6)$$

Where θ_v denote the parameters of the value network, and $V_{\theta_v}(s_t(k))$ is the predicted value of the value network. $\theta_v^{'}$ denote the parameters of the target value network, and $V_{tg}(s_{t+1}(k); \theta_v^{'}, \pi)$ is the target value, which consists of the immediate reward and discounted estimated value of next state, as shown in Equation 7.

$$V_{tg}\left(s_{t+1}; \theta_v^{'}, \pi\right) = \sum_{a_t(k)} \pi\left(a_t(k)|s_t(k)\right) \left(r_{t+1}(k) + \beta V_{\theta_v^{'}}(s_{t+1}(k)) \right) \quad (7)$$

The parameters of the policy network θ_p are updated by the gradient descent rule $\theta_p \leftarrow \theta_p + \lambda_1 \nabla_{\theta_p} \mathcal{L}(\theta_p)$, where λ_1 is the learning rate of the actor, and the gradient is given by Equation 8.

$$\nabla_{\theta_p} \mathcal{L}(\theta_p) = \sum_t \nabla_{\theta_p} \log \pi_{\theta_p}(s_t, a_t) \left(r_{t+1}(k) + \beta V_{\theta_v^{'}}\left(s_{t+1}(k)\right) - V_{\theta_v}(s_t(k)) \right) \quad (8)$$

Since the value function has high variability, we define the advantage function [19] to address it, which is given in Equation 9.

$$A(s_t(k), a_t(k)) = Q(s_t(k), a_t(k)) - V_{\theta_v}(s_t(k)) \quad (9)$$

Where $Q(s_t(k), a_t(k))$ is the Q value for action a_t in state s_t. $V_{\theta_v}(s_t(k))$ is the average value of that state, so this function tells us the improvement compared to the average the action taken at that state. $A(s_t(k), a_t(k)) > 0$ means the gradient is pushed in that direction, and $A(s_t(k), a_t(k)) < 0$ means the action does worse than the average value of that state.

Since

$$Q(s_t(k), a_t(k)) = r_{t+1}(k) + \beta V_{\theta_v^{'}}(s_{t+1}(k)) \quad (10)$$

Combining Equation 9 with Equation 10, we obtain the following Equation 11, which is equivalent to the Temporal-Difference (TD) error [20], so we can use the TD error as an estimation of the advantage function.

$$A(s_t(k), a_t(k)) = r_{t+1}(k) + \beta V_{\theta_v^{'}}(s_{t+1}(k)) - V_{\theta_v}(s_t(k)) \quad (11)$$

The details of the CMA2C are shown in Algorithm 1.

Algorithm 1: CMA2C for E-Taxi Displacement

```
Initialize the value network by randomly selecting parameter θ_v
for i =1 to the maximum iteration number do
    Reset the environment and obtain the initial joint states s_0
    for each time slot t ∈ [0, T) do
        for k=1 to N_t do
            Sample action of each active e-taxi a_t(k) given s_t(k) according to the action probability;
            Compute target value network V_tg by Equation 7 and advantage function A(s_t(k), a_t(k)) by Equation 11 for the policy network, and store the transitions of all active e-taxis (s_t(k), a_t(k), s_{t+1}(k)).
    for j =1 to a certain iteration number M do
        Sample a batch of s_t(k) and V_tg, and then update parameter of the value network θ_v by minimizing the value loss function L(θ_v) over the batch in Equation 6.
        Compute the advantage function A(s_t(k), a_t(k)) by Equation 11 and update the parameter of the policy network θ_p by θ_p ← θ_p + λ_1 ∇_{θ_p} L(θ_p), where ∇_{θ_p} L(θ_p) is calculated according to Equation 8.
```

IV. Evaluation

A. Experimental Setup

Evaluation Data: To evaluate the effectiveness of our displacement system, one-month real-world data collected from the e-taxi fleet in the Chinese city Shenzhen during December 2019 is utilized in this part. During this period, all active taxis are e-taxis, including 20,130 e-taxis, which institute the largest full e-taxi fleet in the world. The one-month e-taxi data include 2.48 billion GPS records and 23.2 million trip records. In addition to e-taxi data, the evaluation dataset also includes the metadata of 123 e-taxi charging stations with over 5,000 charging points. The detailed data formats have been introduced in Section II.

Baseline Setting: We compare our CMA2C-based `FairMove` to the following baselines.

- **GT** is the **G**round **T**ruth, which is obtained from our real-world data. Based on our data, we have historical passenger distributions. We also inferred the charging events of the e-taxis according to the method in [16], and we calculated their cruise time, idle time, profit efficiency, and profit fairness, etc.
- **SD2** is the **S**hortest **D**istance based **D**isplacement [21]. In this setting, the e-taxis are always displaced to serve their nearest passengers or charge in the nearest charging stations no matter what regions the passengers and charging stations are, and it does not have a learning process for a long-term reward. Even though it is a naive method, it is very easy to implement in complicated real-world scenarios. One potential drawback is that some charging stations will be overcrowded in some time slots with this displacement method.
- **TQL** is the standard **T**abular **Q**-**L**earning [22], which is a widely used method for single-agent scenario. It estimates the expected total discounted rewards of state-action pairs by learning a Q-function table with ϵ-greedy policy.
- **DQN** (**D**eep **Q**-**N**etwork) [23] is a popular method in reinforcement learning and has been previously applied to multi-agent settings. DQN learns the action-value function Q^* corresponding to the optimal policy by min-

imizing the loss: $\mathcal{L}(\theta) = \mathbb{E}_{s,a,r,s'}\left[(Q^*(s,a|\theta) - y)^2\right]$, where $y = r + \beta \max_{a'} \bar{Q}^*(s', a')$. where $\bar{Q}$ is a target Q function whose parameters are periodically updated with the most recent θ, which helps stabilize learning.

- **TBA** is the **T**rip **B**andit **A**pproach [6], which is also a reinforcement learning-based method. It is proposed in SIGSPATIAL Cup 2019. It adopts the REINFORCE rule [24] to update the policy. In this setting, e-taxis only know their own states and cannot communicate with each other, so they are purely competitive, and e-taxis will also be displaced to serve their nearest passengers before orders expire. They will also be displaced to charge in the nearest charging stations if they need to charge.

Parameter Setting: The batch size of all deep learning networks is set to be 3500, and we utilize AdamOptimizer with a learning rate of 0.001. We set 10 minutes as a time slot, which is widely adopted by existing works [22], [25], so the one day is divided into $T = 144$ time slots. For the discount factor, we select $\beta = 0.9$ to guarantee the convergence. We set the weighted factor $\alpha = 0.6$ for the following experiments, and we will show the reason in Section IV-B5. All the experiments are repeated 10 times to ensure the robustness of the results.

Evaluation Metrics: Our `FairMove` aims to improve the *profit efficiency* and *profit fairness* of an e-taxi fleet at the same time. According to Equation 2, an implicit indicator of improving e-taxis' the *profit efficiency* is the reduction of their total cruise time T_{cruise}, idle time T_{idle}, or charging duration in electricity pricing peak hours. Hence, we utilize the following metrics to measure the system performance, including (i) *Percentage Reduction of Cruise Time (PRCT)*, (ii) *Percentage Reduction of Idle Time (PRIT)*, (iii) *Percentage Increase of Profit Efficiency (PIPE)*, and (iv) *Percentage Increase of Profit Fairness (PIPF)*. In addition, we also show the impact of the parameter α on the system performance.

$$PRCT(D) = \frac{\sum_{i=1}^{M} T_{cruise}^{(i)}(G) - \sum_{i=1}^{M} T_{cruise}^{(i)}(D)}{\sum_{i=1}^{M} T_{cruise}^{(i)}(G)} \times 100\% \quad (12)$$

$$PRIT(D) = \frac{\sum_{j=1}^{Z} T_{idle}^{(j)}(G) - \sum_{j=1}^{Z} T_{idle}^{(j)}(D)}{\sum_{j=1}^{Z} T_{idle}^{(j)}(G)} \times 100\% \quad (13)$$

where $T_{cruise}^{(i)}(D)$ is the *cruise time* for ith trip under the displacement strategy D, which could be SD2, TQL, DQN, or `FairMove` (based on CMA2C); M is the total number of trips served by the e-taxi fleet. $T_{cruise}^{(i)}(G)$ is the *cruise time* for ith trip of the Ground Truth; $T_{idle}^{(j)}(D)$ is *idle time* of the jth charging events under displacement strategy D; Z is the total number of charging events of the e-taxi fleet; $T_{idle}^{(j)}(G)$ is the *idle time* of the jth charging events of the Ground Truth; N is the total number of e-taxis in the fleet; The Ground Truth is obtained by merging the GPS data, transaction data, and the charging station data.

$$PIPE(D) = \frac{\sum_{k=1}^{N} PE_k(D) - \sum_{k=1}^{N} PE_k(G)}{\sum_{k=1}^{N} PE_k(G)} \times 100\% \quad (14)$$

$$PIPF(D) = \frac{PF(G) - PF(D)}{PF(G)} \times 100\% \quad (15)$$

where $PE_k(D)$ is the *profit efficiency* of the e-taxi k under the displacement strategy D, which can be SD2, TQL, DQN, or `FairMove` (based on CMA2C); $PE_k(G)$ is the *profit efficiency* of the e-taxi k without any external displacement; $PF(G)$ is the *profit fairness* of the Ground Truth; $PF(D)$ is profit fairness of the displacement strategy D.

B. Displacement Performance

1) Cruise Time Comparison: Since a key impact factor of profit efficiency of e-taxis is their *cruise time*, we compare our `FairMove` to other state-of-the-art baselines considering the cruise time reduction. Fig. 10 shows the each cruise time distribution under different displacement methods. We found all methods reduce the cruise time for seeking passengers at different degrees compared to the ground truth due to their centralized management mode. The median value of the cruise time without other displacement is around 6.5 minutes, and it decreases to 5.4 minutes under our `FairMove` displacement. In addition to the decrease of the median value of the cruise time, its variance also becomes smaller with `FairMove` displacement, which could be induced by our fairness consideration. Fig. 11 shows the average PRCT for all trips during different hours of a day. We found our `FairMove` achieves the best performance compared to other methods. Particularly, `FairMove` reduces over 40% of cruise time for e-taxis during 5:00-7:00, when there are few passenger demands and drivers need to cruise a longer time to find a passenger without centralized displacement.

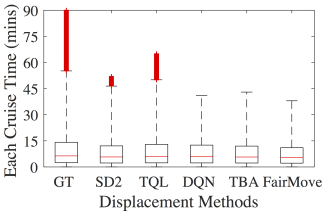

Fig. 10. Per-trip cruise time for seeking passengers.

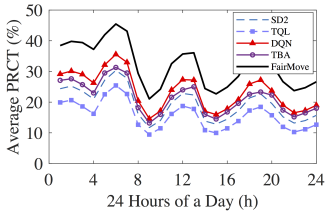

Fig. 11. Average PRCT distribution in different hours.

In general, our `FairMove` achieves 32.1% of PRCT for each trip compared to the ground truth on average, as shown in Table II. The reason could be that deep reinforcement learning-based `FairMove` not only considers the short-term immediate benefits but also considers the long-term benefits. DQN also has a good performance with 23.6% of PRCT, followed by TBA and SD2 with 21.3% and 19.4% of PRCT compared to the ground truth.

TABLE II
AVERAGE PERCENTAGE REDUCTION OF CRUISE TIME (PRCT).

Methods	SD2	TQL	DQN	TBA	FairMove
PRCT	19.4%	15.7%	23.6%	21.3%	32.1%

2) Idle Time Comparison: Since the idle time for charging will also impact the profit efficiency of the e-taxis, we also compare our FairMove to other state-of-the-art baselines considering the idle time reduction. Fig. 12 shows the idle time distribution for each charging event under different displacement methods. We found that our FairMove achieves the best performance, and 75% of the per-charge idle time is less than 22 minutes. However, SD2 prolongs the idle time since many e-taxis around charging stations will be displaced to the same charging stations, which causes long queuing in the overcrowded charging stations. Fig. 13 shows the average PRIT of all charging events during 24 hours of a day. We found our FairMove achieves the most PRIT during the high charging demand hours, e.g., 4:00-5:00 and 17:00-18:00, which potentially indicates our method can also benefit the charging issues for e-taxis, especially for addressing the intensive charging peaks.

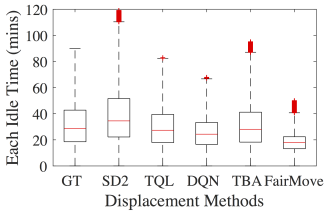

Fig. 12. Per-charge idle time distribution.

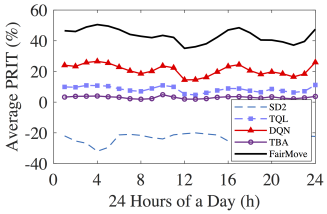

Fig. 13. Average PRIT distribution in different hours.

In general, our FairMove achieves 43.3% of PRIT for each charging event compared to the ground truth on average, as shown in Table III. The reason would be that deep learning-based FairMove will choose the stations with the consideration of long-term benefits. DQN also has a good performance with 21% of PRIT. However, SD2 has a negative PRIT, which means it prolongs the idle time as many near e-taxis have been displaced to the same charging stations, resulting in long queuing in these stations. Although TBA may also cause some charging stations overcrowded, it achieves 3.1% of PRIT due to the long-term benefit consideration and potential cruise time reduction with the reinforcement learning method.

TABLE III
AVERAGE PERCENTAGE REDUCTION OF IDLE TIME (PRIT).

Methods	SD2	TQL	DQN	TBA	FairMove
PRIT	-23.1%	8.4%	21%	3.1%	43.3%

3) Profit Efficiency Comparison: Since one of the most important objectives of the paper is to improve the profit efficiency of the e-taxi fleet, we compare FairMove to baselines considering their profit efficiency changes. Fig. 14 shows the hourly profit efficiency of each e-taxi under different displacement methods. We found the hourly profit efficiency varies from 0 to 120 without displacement, and the median value is 45.2. The profit efficiency of SD2 has a slight decrease due to the prolonged idle time. Both TQL and DQN increase the hourly profit efficiency for e-taxis on average, but our FairMove achieves the best performance, with the median value of 53.1. In addition, the variance between the e-taxis becomes smaller since we consider the fairness between them.

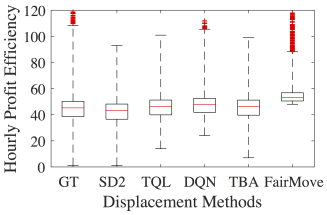

Fig. 14. Hourly profit efficiency.

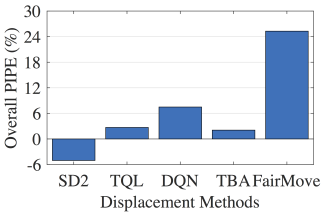

Fig. 15. Overall PIPE comparison.

In Fig. 15, we show the overall PIPE in one month of different displacement methods. We found our FairMove increases the profit efficiency for the e-taxi fleet by 25.2%, followed by DQN with a 7.5% of increase. However, SD2 reduces the profit efficiency for the e-taxi fleet by 5% due to the prolonged idle time.

4) Profit Fairness Comparison: Another key objective of the paper is to improve the profit fairness for the e-taxi fleets. From 16, we found our FairMove achieves the best performance with 54.7% of PIPF. The reason may be that we formulate the problem with fairness as a part of the objective function, and solve it by deep reinforcement learning methods, so it not only improves the profit efficiency but also improves the profit fairness for the e-taxi fleet. SD2 and TBA achieve similar improvement of the profit fairness of the e-taxi fleet by 13%. Due to the fairness consideration, TQL and DQN also improve the fairness efficiency by 28.7% and 17.9%, respectively.

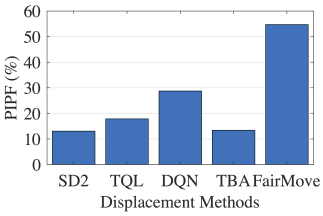

Fig. 16. PIPF comparison.

5) Performance Under Different Weighted Factor α: In this subsection, we conduct a sensitivity analysis to study the impacts of the reward weighted factor α for training of the proposed multi-agent deep reinforcement learning method. As mentioned, α measures the tradeoff between the overall profit efficiency of the fleet and the fairness between individual e-taxis. The higher the α, the more emphasis on the overall profit efficiency. The lower the α, the more emphasis on fairness. We compare the performance of the FairMove with different weighted factor α (from 0 to 1 with a step of 0.2). The average reward r of the proposed CMA2C is shown in Table IV, which shows that setting the parameter α from 0.6 to 0.8 leads to the best system performance. Since maximizing fairness alone may harm the overall profit efficiency for the e-taxi fleet, this finding is reasonable. This is also the reason why we select $\alpha = 0.6$ for the above comparisons.

TABLE IV
SYSTEM PERFORMANCE UNDER DIFFERENT WEIGHTED FACTOR α.

Weight Factor α	0	0.2	0.4	0.6	0.8	1
Average Reward r	6.95	7.05	7.16	7.44	7.39	7.15

V. Discussions

Data-Driven Findings. Based on our data-driven investigation, we obtain some new findings. (i) Idle time reduction does not necessarily indicate the prolonged time for serving passengers since some drivers may need to spend more time to seek passengers after charging in some regions with low passenger travel demand (Fig. 5 and Fig. 6), which is rarely considered by existing charging recommendation works. (ii) The potential revenue for serving passengers after charging in different stations may be also different, which is highly dynamic in both spatial and temporal dimensions (Fig. 7).

Generalization on Electric Ridesharing Fleets. Even though the setting of this paper is for e-taxis, we believe it has the potential to be generalized to electric ridesharing fleets, e.g., Uber, Lyft, and DiDi. For the ridesharing fleets, not only vehicles' real-time information is uploaded, but also the passengers' request information will be recorded, so it is easy to obtain passengers' origins & destinations for more accurate displacement decisions. In addition, ridesharing fleets are already under a centralized management mode, so it is feasible to apply our displacement system to them. In the future, we will also verify the performance of our fairness-aware displacement system on electric ridesharing fleets.

Fairness of Different Driver Groups. Since drivers may have different performance, which is decided by many factors like taxi driving years, accidents, and reputation, it is also reasonable to divide all drivers into different groups by their performance levels and quantify their fairness within the same group. Even though we did not divide the drivers into different groups, we found the government and taxi companies have already comprehensively evaluated each driver's performance based on multiple factors and label it on the taxi [26], which is normally represented by a five-star rating. Hence, we can directly merge it into our displacement system for five groups and achieve fairness in the same group.

VI. Related Work

A. Traditional Gas Taxi Dispatching

In the last decade, with the wide development of mobile sensors and advanced data processing technologies, a large number of works have been done to improve the service efficiency of taxi fleets based on real-world data, e.g., taxi GPS data and transaction data. [6], [7] in the SIGSPATIAL Cup 2019 tried to optimize the efficiency of taxi drivers searching for customers, and they considered all taxis as the competitive relationship, which potentially causes inequality between taxis. Although [6] adopted a reinforcement learning-based method for vehicle displacement, it only utilized the traditional REINFORCE rule [24] to update the policy, while we utilized two networks (i.e., Critic and Actor) parameterized with deep neural networks to learn the optimal policy, which has the capability to approximate parameters for complicated tasks in a highly dynamic environment. Our critic-actor structured reinforcement learning considers the cooperation of agents to achieve a fairness-aware solution, while agents in [24] are purely competitive. Furthermore, our method is in a centralized training and decentralized execution fashion, which makes it more efficient for decision making. In addition, the charging issues cannot be naively addressed by the solutions of [6], [7] due to many practical issues, e.g., possible overcrowded charging stations and time-variant charging pricing. Miao et al. [27] presented a receding horizon control framework to dispatch taxis, which applied predicted models and sensing data to decide dispatch locations for vacant taxis considering different objectives, e.g., reducing the average total idle distance and supply-demand ratio error. Yuan et al. [28] presented a recommendation system to help taxi drivers to pick up passengers quickly and maximize the profit of the next trip. However, all these works failed to consider the fairness between taxi drivers when they make dispatching decisions. Even though [29] proposed a route assignment mechanism for fair taxi route recommendations, it focused on the conventional gas taxis, which has different operation patterns and energy replenishment mechanisms with e-taxis. In addition, the complicated charging process has not been considered, which makes it challenging to be reapplied for e-taxi displacement.

B. Electric Taxi Charging Scheduling

With the rapid taxi electrification process, more and more research [4], [9], [16], [30], [31] focuses on e-taxi charging issues. Among all these works, e-taxi charging scheduling is one of the most popular topics. Dong et al. [9] developed a real-time charging scheduling framework for e-taxi fleets to reduce the queuing time of e-taxis. Yuan et al. [30] proposed a charging scheduling framework for e-taxis to minimize the traveling time to charging stations and waiting time at charging stations with considering the dynamic passenger demand. However, all these works only focused on the charging issues of e-taxis without considering the potential revenue loss related to charging. In addition, they neglected the fairness between taxis, which may potentially cause drivers not to follow their scheduling decisions and make the system unsustainable.

Recently, there are some works [16], [32] trying to seek fairness-aware scheduling for e-taxis. Yang et al. [32] proposed a charging coordination solution for e-taxis to reduce their queuing time in charging stations. Wang et al. [16] designed a fairness-aware Pareto efficient charging recommendation system called FairCharge to minimize the total charging idle time (traveling time + queuing time) in charging stations combined with fairness constraints. However, all these works only considered the charging processes of e-taxis while neglected their overall revenue, which is a key concern of taxi drivers.

C. Uniqueness of Our Work

In summary: To our best knowledge, our `FairMove` is the first displacement system for e-taxis to improve the overall profit efficiency of taxi fleets with the fairness consideration. `FairMove` considers not only the passenger demand but also the complicated charging issues of e-taxis (e.g., unique charging behaviors and time-varying electricity pricing). Moreover, `FairMove` also emphasizes the fairness between e-taxis.

VII. Conclusion

In this paper, we design the first data-driven fairness-aware displacement system called `FairMove` based on multi-source data, which aims to improve the overall profit efficiency and profit fairness of the entire e-taxi fleet in a city. We first conduct a data-driven investigation, which shows the uniqueness of the e-taxi displacement problem. We then formulate the e-taxi displacement as a multi-agent deep reinforcement learning and propose a centralized multi-agent actor-critic (CMA2C) approach to tackle this problem. `FairMove` considers not only dynamic passenger demand & supply in both temporal and spatial dimensions but also considers the complicated charging problems (e.g., time-variant electricity pricing) and per-trip profit. We implement and evaluate our `FairMove` based on a real-world dataset obtained from the large-scale full e-taxi fleet including over 20,100 vehicles. The extensive experimental results show that our fairness-aware `FairMove` effectively improves the profit efficiency and profit fairness by 25.2% and 54.7%, respectively.

Acknowledgment

The authors would like to thank anonymous reviewers for their valuable comments and suggestions. This work is partially supported by NSF 1849238, NSF 1932223, NSF 1951890, NSF 1952096, NSF 2003874.

References

[1] Y. Li, J. Luo, C.-Y. Chow, K.-L. Chan, Y. Ding, and F. Zhang, "Growing the charging station network for electric vehicles with trajectory data analytics," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015, pp. 1376–1387.

[2] V. I. Roland Irle, Jos Pontes, "The electric vehicle world sales database," http://www.ev-volumes.com/, 2018.

[3] S. Ma, Y. Zheng, and O. Wolfson, "T-share: A large-scale dynamic taxi ridesharing service," in *2013 IEEE 29th International Conference on Data Engineering (ICDE)*. IEEE, 2013, pp. 410–421.

[4] G. Wang, W. Li, J. Zhang, Y. Ge, Z. Fu, F. Zhang, Y. Wang, and D. Zhang, "sharedcharging: Data-driven shared charging for large-scale heterogeneous electric vehicle fleets," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–25, 2019.

[5] X. Zhou, H. Rong, C. Yang, Q. Zhang, A. V. Khezerlou, H. Zheng, M. Z. Shafiq, and A. X. Liu, "Optimizing taxi driver profit efficiency: A spatial network-based markov decision process approach," *IEEE Transactions on Big Data*, 2018.

[6] F. Borutta, S. Schmoll, and S. Friedl, "Optimizing the spatio-temporal resource search problem with reinforcement learning (gis cup)," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, pp. 628–631.

[7] J.-S. Kim, D. Pfoser, and A. Züfle, "Distance-aware competitive spatiotemporal searching using spatiotemporal resource matrix factorization (gis cup)," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, pp. 624–627.

[8] G. Wang, F. Zhang, H. Sun, Y. Wang, and D. Zhang, "Understanding the long-term evolution of electric taxi networks: A longitudinal measurement study on mobility and charging patterns," *ACM Transactions on Intelligent Systems and Technology (TIST)*.

[9] Z. Dong, C. Liu, Y. Li, J. Bao, Y. Gu, and T. He, "Rec: Predictable charging scheduling for electric taxi fleets," in *Real-Time Systems Symposium (RTSS), 2017 IEEE*. IEEE, 2017, pp. 287–296.

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[11] Wikipedia, "Byd e6," https://en.wikipedia.org/wiki/BYD_e6, 2020.

[12] G. electric vehicle, "Shenzhen will promote over 10,000 electric taxis. how to deal with the complicated charging issues?" https://www.gg-ev.com/asdisp2-65b095fb-23169-.html, 2020.

[13] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[14] C. H. Liu, Y. Zhao, Z. Dai, Y. Yuan, G. Wang, D. Wu, and K. K. Leung, "Curiosity-driven energy-efficient worker scheduling in vehicular crowdsourcing: A deep reinforcement learning approach," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 25–36.

[15] M. Fleurbaey and F. Maniquet, *A theory of fairness and social welfare*. Cambridge University Press, 2011, vol. 48.

[16] G. Wang, Y. Zhang, Z. Fang, S. Wang, F. Zhang, and D. Zhang, "Faircharge: A data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–25, 2020.

[17] K. Lin, R. Zhao, Z. Xu, and J. Zhou, "Efficient large-scale fleet management via multi-agent deep reinforcement learning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1774–1783.

[18] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[19] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[20] H. Van Seijen, A. R. Mahmood, P. M. Pilarski, M. C. Machado, and R. S. Sutton, "True online temporal-difference learning," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5057–5096, 2016.

[21] F. Stollar, J. McCann, and R. Chatley, "Fleet management in on-demand transportation networks: Using a greedy approach," *Imperial College London*, 2018.

[22] I. Jindal, Z. T. Qin, X. Chen, M. Nokleby, and J. Ye, "Optimizing taxi carpool policies via reinforcement learning and spatio-temporal mining," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1417–1426.

[23] J. Foerster, I. A. Assael, N. De Freitas, and S. Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," in *Advances in neural infor. processing systems*, 2016, pp. 2137–2145.

[24] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine learning*, vol. 8, no. 3-4, pp. 229–256, 1992.

[25] J. Jin, M. Zhou, W. Zhang, M. Li, Z. Guo, Z. Qin, Y. Jiao, X. Tang, C. Wang, J. Wang *et al.*, "Coride: Joint order dispatching and fleet management for multi-scale ride-hailing platforms," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1983–1992.

[26] I. News, "Star rating of shenzhen taxi drivers," http://news.ifeng.com/gundong/detail_2014_01/14/33002787_0.shtml, 2014.

[27] F. Miao, S. Han, S. Lin, J. A. Stankovic, D. Zhang, S. Munir, H. Huang, T. He, and G. J. Pappas, "Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 463–478, 2016.

[28] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 10, pp. 2390–2403, 2012.

[29] S. Qian, J. Cao, F. L. Mouël, I. Sahel, and M. Li, "Scram: A sharing considered route assignment mechanism for fair taxi route recommendations," in *Proceedings of the 21th ACM SIGKDD*, 2015, pp. 955–964.

[30] Y. Yuan, D. Zhang, F. Miao, J. Chen, T. He, and S. Lin, "p^ 2charging: Proactive partial charging for electric taxi systems," in *2019 IEEE 39th ICDCS*. IEEE, 2019, pp. 688–699.

[31] G. Wang, X. Chen, F. Zhang, Y. Wang, and D. Zhang, "Experience: Understanding long-term evolving patterns of shared electric vehicle networks," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–12.

[32] Z. Yang, T. Guo, P. You, Y. Hou, and S. J. Qin, "Distributed approach for temporal–spatial charging coordination of plug-in electric taxi fleet," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3185–3195, 2018.