Mobility and Energy Management in Electric Vehicle Based Mobility-on-Demand Systems: Models and Solutions

Liang Ni, Bo Sun, *Member, IEEE*, Xiaoqi Tan, and Danny H. K. Tsang, *Life Fellow, IEEE*

***Abstract*— An electric vehicle based mobility-on-demand (EMoD) system provides shared transportation (e.g., car-sharing or ride-sharing) to satisfy customers' individual mobility demands. It has been recognized as a vital alternative form of transportation between public and private transportations in future sustainable cities. Constrained by the long charging time and limited driving range of EVs, an operator of an EMoD system demands for decision-making models and algorithms to manage the mobility and energy of EVs to best serve customers with least costs. In this paper, we propose a stochastic dynamic program (DP) to model three operational decisions of the EMoD system: i) dispatching EVs to serve mobility demand from customers, ii) repositioning EVs to accommodate the unbalanced mobility demands between service regions, and iii) recharging EVs to maintain their sufficient state-of-charge levels. To handle this large-scale DP problem, we first observe and prove that it has a coordinate-wise concave value function. Based on this structural property, we propose to use a separable piecewise linear function to approximate the value function and design an approximation-based algorithm to efficiently derive the decision policy. Numerical tests show that our proposed algorithm significantly outperforms the existing model-free approaches (e.g., greedy heuristic and Q-learning) that fail to take into account the structural properties of the DP problem.**

***Index Terms*— Mobility-on-demand, sequential decision-making, dynamic programming, perturbation analysis.**

I. Introduction

THE world is undergoing a process of rapid urbanization. It is estimated that more than two-thirds of the world's population will live in urban areas by 2050 [1]. However, such urbanization creates enormous problems of environmental pollution and decreases the general quality of life of urban residents [2]. Driven by these environmental and societal challenges, urban mobility systems are experiencing a dramatic transformation to become more sustainable and cost effective. In conventional urban mobility systems, public transportation cannot satisfy individual mobility services well, and thus urban residents commonly choose to purchase private cars for private transportation. However, private cars are typically parked for more than 90% of the time, but contribute substantially to environmental pollution and traffic congestion [3]. Recently, mobility-on-demand (MoD) systems have emerged as a vital alternative for public and private transportations by providing shared transportation tools (e.g., bikes, cars) to satisfy customers' individual mobility demands [4], [5]. For example, car-sharing (e.g., Zipcar) and ride-sharing (e.g., Uber) based MoD platforms have been proved to be successful business models [6]. Moreover, since electric vehicles (EVs) are more energy efficient and almost zero emission, they are widely accepted as substitute to gasoline vehicles. With the electrification of transportation, it is envisioned that an EV based MoD (EMoD) system will become an inevitable trend in future sustainable cities [7].

The core operational problem of the current MoD system is how to dynamically reposition gasoline vehicles across service regions to meet the mobility demand that varies in both time and space [8], [9], [10]. However, due to the long charging time and limited driving range of EVs, an EMoD system has to address new problems deriving from the inevitable energy management of EVs. First, an efficient recharging scheme is necessary to maintain enough energy for EVs to provide mobility services with guaranteed quality-of-service[1] (QoS). Second, a joint management of mobility and energy is needed to reduce total recharging cost due to the temporal and spatial difference in recharging opportunities (e.g., location- and time-dependent electricity prices). Third, an EMoD system can offer undifferentiated services to a mobility demand using any EV with sufficient energy. However, EVs with different energy levels may affect the recharging costs and the probability of successfully satisfying future mobility demands. Thus, the

Manuscript received 15 August 2020; revised 29 December 2021 and 21 July 2022; accepted 12 December 2022. This work was supported by the Hong Kong Research Grant Council (RGC) General Research Fund through "Mobility and Energy Management of Electric Vehicle-Based Mobility-on-Demand Systems in Future Smart Cities" under Project 16202619. The Associate Editor for this article was A. Bucchiarone. *(Corresponding author: Bo Sun.)*

Liang Ni is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: lniaa@connect.ust.hk).

Bo Sun is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: bsun@cse.cuhk.edu.hk).

Xiaoqi Tan is with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mail: xiaoqi.tan@ualberta.ca).

Danny H. K. Tsang is with The Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou, Guangdong 511400, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: eetsang@ust.hk).

Digital Object Identifier 10.1109/TITS.2022.3231435

[1]QoS refers to the capability of a system to provide better services. For instance, we can define the QoS of an EMoD system as the probability of providing available mobility services upon requests.

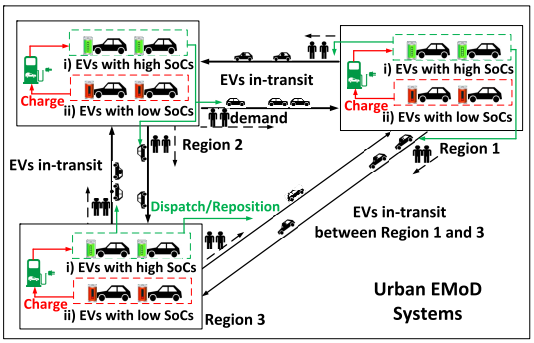


Fig. 1. An illustration of urban EMoD systems. A fleet of EVs are owned and operated by the EMoD system operator. EVs with sufficiently high SoCs can be dispatched to serve mobility demand or repositioned to other regions (green arrow). After being dispatched or repositioned through travel links between regions, EVs will travel for a certain time and consume a certain amount of energy before arriving in the destination region. The EVs with insufficient SoCs need to be recharged before traveling to other regions (red arrow).

operator of an EMoD system needs to decide which EV with a particular state-of-charge (SoC) level should be dispatched to serve which mobility demand.

The mobility and energy management of EMoD system involves a complicated sequential decision-making process. To address this problem, we aim to make three types of decisions to maximize the overall revenue of the EMoD system: i) dispatching EVs (i.e., deciding which of the available EVs) to serve mobility demand with QoS guarantee, ii) repositioning the EVs that are currently not serving customers among service regions to accommodate the unbalanced mobility demand, and iii) recharging EVs with low SoCs to ease the range-anxiety issue. We illustrate the EMoD system and its operational decisions in Fig. 1. To capture such sequential decision-making process, we formulate a stochastic dynamic program (DP) problem. However, it is computationally challenging to derive an optimal policy in a practical-scale EMoD system, faced with the non-stationary mobility demand and charger availability (number of available chargers). To develop a computationally tractable policy, we utilize approximate dynamic programming (ADP) techniques [11], [12]. Based on the rigorously proved coordinate-wise concave property of the value function in our DP formulation, we use a separable piecewise linear function to approximate the value function, and design an ADP-based algorithm to make decisions efficiently. Before introducing the technical details of the EMoD system model and the decision strategy, we present the related literature in the following section.

A. Related Work

A rich literature on shared transportation has emerged in recent years, where the mobility management problem in gasoline vehicle-based MoD systems is extensively investigated. The common objective is to design a vehicle repositioning strategy so that the total repositioning cost is minimized [8], [9], [13], or the system revenue can be maximized [10]. Specifically, [13] formulates the vehicle repositioning problem in a vehicle sharing system into a stochastic DP. By dynamically matching the vehicle supply to the travel demand, it can achieve the minimal expected total repositioning cost.

Different from the gasoline vehicle-based MoD systems, the EMoD system needs a joint management of mobility and energy, which is less studied in the literature. The authors of [14] and [15] investigate the optimal routing of EVs with in-route charging to serve mobility demand in an EMoD system. Based on the global information of charging station availability and electricity prices, the total energy cost is minimized in [14]. Leveraging the probabilistic information of customers' arrivals, the overall trip time for customers is minimized in [15]. In [16], a queuing model is introduced by assuming the arrival processes of mobility demand and available EVs as Poisson processes. Planning decisions on proportions of vehicles to be charged or serve mobility demands are derived under stability conditions, such that the total response time of the system can be minimized. Reference [17] focuses on making planning decisions on the size of EV fleet and the number of charging facilities in a national scale EMoD system, based on the estimation of daily customers' demand. A quadratic programming problem is formulated to minimize the total system cost. In [18], a max-flow problem is solved to predetermine a set of feasible trips given the available EVs at each location in an EMoD system, where customers announce their demands in a day-ahead manner. Although these results are appropriate for planning problems of EMoD systems, the applicability of real-time decisions, such as recharging and repositioning facing uncertain mobility demand, needs further investigation.

To capture the uncertain mobility demand, another stream of works [19], [20], [21], [22] uses agent-based simulations to evaluate the environmental and economic impacts of shared autonomous EVs. However, there is still a lack of a model-based mathematical framework to make operational decisions for EMoD systems. The authors of [23] consider an online charging scheduling problem of the autonomous EVs in an EMoD system. To handle the uncertain daily sequence of mobility demand, an online algorithm is designed to maximize the system welfare, with a worst-case performance guarantee. However, with the main focus on charging scheduling, vehicle repositioning is neglected. To the best of our knowledge, our paper is the first addressing the sequential decision-making model on both mobility and energy management in an EMoD system, which makes our problem more challenging and differentiates our work from existing papers.

B. Our Contribution

Our main contributions are summarized as follows:

- We consider the joint mobility and energy management problem in the EMoD system. Instead of making planning decisions, we design a sequential decision-making model to make real-time decisions in a practical-scale EMoD system, with non-stationary mobility demand and charger availability.
- We formulate the mobility and demand management problem into a stochastic DP problem, which is computationally challenging to derive the optimal policy. Different

TABLE I
PARAMETERS AND VARIABLES

Notation	Definition
$\mathcal{M}, M$	The set and the total number of service regions
$\mathcal{L}, L$	The set and the total number of paths
i, i^ℓ	The region index and the origin region of path $\ell \in \mathcal{L}$
$\mathcal{T}, T$	The set and the total number of decision horizons
$\mathcal{K}, K$	The set and the total number of SoC levels
$\mathcal{K}^\ell$	The set of SoC levels, which are sufficient to travel on ℓ
ξ_t^ℓ	The demand at time epoch t on path ℓ
ζ_t^i	The number of chargers in region i at time epoch t
N	Total number of EVs in the system
v_t^{ik}	The number of EVs with SoC k in region i at time epoch t
$\mu_t^{\ell k}$	The number of EVs starting to travel at time epoch t with an SoC k on path ℓ
$\mathcal{I}_t, \mathcal{P}_t$	The information and physical space of time period t
r_t^ℓ, η_t^i	The reward and penalty for serving and losing one customer in time period t
$\gamma_t^\ell, \lambda_t^\ell$	Per-EV repositioning and recharging costs in period t.
$\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t$	The vectors of repositioning, recharging and dispatching decision variables
$\boldsymbol{s}_t, \boldsymbol{a}_t$	The vector of system state and action of time period t
$V_t, \tilde{V}_t, \hat{V}_t, \bar{V}_t$	The exact value function, the continuous extension of V_t, the approximate value function and the estimation of V_t from problem (14)
$c, \mathcal{C}$	The information interval index and the set of information intervals
$d, \mathcal{D}$	The coordinate index and the coordinate index set of the physical state
$g_{d,t}^c, \theta_{d,t}^c$	The approximate function and slope value for information interval c, coordinate d of time period t

from existing model-free methods (e.g., Q-learning), we rigorously prove that the value function of the DP problem is coordinate-wise concave. By leveraging such structural property, we approximate the value function as a separable piecewise linear function and design an ADP-based algorithm to make near-optimal decisions efficiently.

- We perform extensive numerical tests based on a Hong Kong transportation system. Compared with heuristic and model-free algorithms, our proposed algorithm can achieve significant performance improvement and meanwhile meet the time requirement for making real-time decisions.

II. EMoD System Model and DP Formulation

In this section, we present the detailed EMoD system model in Sec. II-A and the joint mobility and energy management problem in Sec. II-B. In Table I, we summarize all parameters and variables in our model.

A. EMoD System Model

We consider an EMoD system that owns and operates a fleet of N EVs to provide on-demand mobility services in M regions. Let $\mathcal{M} := \{1, \ldots, M\}$ and $\mathcal{L} := \{1, \ldots, L\}$ denote the index sets of the service regions and (virtual) travel links between regions, respectively. For any service region $i \in \mathcal{M}$, we denote the set of travel links originating from i by $\mathcal{L}^i \subset \mathcal{L}$. For any link $\ell \in \mathcal{L}$, we denote its origin region by i^ℓ. Each customer arrives and submits its mobility demand to travel on a particular link $\ell \in \mathcal{L}$. The EMoD operator will then dispatch an EV with a certain SoC level in region i^ℓ to pick up the customer and travel through link ℓ. We focus on a discrete-time stochastic system with a finite time horizon of T periods. The operator's decisions are made in discrete time epochs $t \in \mathcal{T} := \{1, \ldots, T\}$, while demand from customers can arrive any time between two epochs.

To serve customers' mobility demand, we need to address the following uncertainties in the EMoD system. First, customers randomly arrive for mobility demand at different time from varying locations. This uncertainty results in the temporal and spatial imbalance of the supply and demand of the EVs. Second, EVs in the EMoD system need to share chargers with other EVs in public charging facilities. Thus, the number of available chargers at different time and locations is also uncertain. We model the mobility demand and the number of available chargers as random variables, which will be explained in detail in Sec. II-B. Also note that the traveling time and energy consumption of EVs on different links depend on time-varying traffic conditions. More congested traffic condition usually leads to longer traveling time and more energy consumption. The uncertainties of traffic conditions are not directly introduced in the system model but we will show how to incorporate this uncertainty in the numerical tests. For detailed numerical settings, please refer to Sec. IV. We would like to point out that, in this paper, we aim to propose a general model and its general solution to capture as many characteristics of the real system as possible. In a more specific case that a fixed number of available chargers are reserved particularly for EMoD systems, and the time and energy consumption on each link are given, our model and algorithm derived in this paper can handle it in a simpler way (with fewer system states in the model, Algorithm 1 introduced later in Sec. III-D can be executed with lower computational burden).

B. Problem Formulation

In order to deal with the uncertain mobility demand and charger availability in the EMoD system, we make sequential decisions, including repositioning, recharging and dispatching, with the purpose of maximizing the overall system revenue of T time periods. To further capture the impact of current decisions on the future, the sequential decision-making process is formulated into a finite time horizon stochastic DP.

1) State Space: The state of the EMoD system can be categorized into information state and physical state.

Suppose an operator owns a group of homogeneous EVs, each of which needs K periods to be fully charged. By setting the (normalized) delivered energy of charging an EV for one time slot as an SoC level, we can denote the set of EVs' SoC levels by $\mathcal{K} := \{0, \ldots, K\}$. Let ξ_t^ℓ and ζ_t^i denote the mobility demand on travel link ℓ that arrives during $(t-1, t)$ and the number of available chargers in region i at time epoch t, respectively. At time epoch t, the system operator can observe both $\boldsymbol{\xi}_t := \{\xi_t^\ell | \ell \in \mathcal{L}\}$ and $\boldsymbol{\zeta}_t := \{\zeta_t^i | i \in \mathcal{M}\}$. We define the information state of the EMoD system as $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ and denote

its feasible space by $\mathcal{I}_t$. As mentioned in Sec. II-A, ξ_t^ℓ and $\zeta_t^i, t \in \mathcal{T}$ are regarded as random variables to generalize our system settings, which are realized at the beginning of each time period t. In more specific settings where a fixed number of chargers is reserved particularly for the EMoD system, the number of available chargers ζ_t^i can be considered as a constant in each period since the charging decisions for all EVs with different SoCs are made in every time period. The charging decision will be further explained in the following paragraphs.

In addition to the information state, the system operator can also observe the physical distribution of EVs in all regions $\boldsymbol{v}_t := (v_t^{1,0}, \ldots, v_t^{i,k}, v_t^{i,k+1}, \ldots, v_t^{M,K})$, where $v_t^{i,k}, i \in \mathcal{M}, k \in \mathcal{K}$ represents the number of EVs with SoC k in region i at time epoch t. Moreover, the number of in-transit EVs, $\boldsymbol{u}_t := (u_t^{1,0}, \ldots, u_t^{\ell,k}, u_t^{\ell,k+1}, \ldots, u_t^{L,K})$, can be observed, where $u_t^{\ell,k}, \ell \in \mathcal{L}, k \in \mathcal{K}$ is the number of EVs starting to travel at time epoch t on link ℓ with original SoC k. Let $\boldsymbol{u}_{1:t-1} := (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{t-1})$ denote the concatenated historical records up to time period $t-1$. The state space of the physical state $(\boldsymbol{v}_t, \boldsymbol{u}_{1:t-1})$ in time period t is denoted by $\mathcal{P}_t$, which has a dimension of $M \times (K+1) + L \times (K+1) \times (t-1)$ based on the above definitions. $\mathcal{P}_t$ is constrained by the total number of EVs in the system:

$$\sum_{i \in \mathcal{M}} \sum_{k \in \mathcal{K}} v_t^{i,k} + \sum_{\ell \in \mathcal{L}} \sum_{k \in \mathcal{K}^\ell} \sum_{\iota = t - \tau^\ell + 1}^{t-1} u_\iota^{\ell,k} = N, v_t^{i,k}, u_\iota^{\ell,k} \in \mathbb{N}. \tag{1}$$

We denote the subset of SoC levels, which are sufficient to travel on link ℓ, by $\mathcal{K}^\ell = \{k | k \geq k^\ell, k \in \mathcal{K}\}$. Additionally, k^ℓ and τ^ℓ represent the energy consumption (SoC levels) and traveling time on link ℓ, respectively. Note that these values are generally stochastic and non-stationary, depending on traffic conditions. $\mathbb{N}$ is used to represent the natural number set. The second term of (1) on the left-hand side indicates the total number of in-transit EVs at time epoch t. Particularly, $\sum_{\iota=t-\tau^\ell+1}^{t-1} u_\iota^{\ell,k}$ is the total number of in-transit EVs that are traveling on link ℓ. Only the EVs that start to arrive after $t - \tau^\ell + 1$ are counted since those earlier ones have already gone through the link. Thus, we can define the system state of EMoD system in time period t (before making operational decisions) as $\boldsymbol{s}_t = \{\boldsymbol{v}_t, \boldsymbol{u}_{1:t-1}, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t\}$, with the state space $\mathcal{S}_t = \mathcal{P}_t \times \mathcal{I}_t$.

2) Action Space: After observing both the information and physical states, the system operator can make the following decisions: i) Reposition EVs between different service regions. We denote the repositioning decision by $\boldsymbol{x}_t = \{x_t^{\ell,k} | \ell \in \mathcal{L}, k \in \mathcal{K}\}$, where $x_t^{\ell,k}$ represents the number of EVs in region i^ℓ with SoC level k that are repositioned through link ℓ at time epoch t. ii) Recharge EVs at available chargers. We denote the recharging decision by $\boldsymbol{y}_t = \{y_t^{i,k} | i \in \mathcal{M}, k \in \mathcal{K}\}$, where $y_t^{i,k}$ is the number of EVs with SoC level k that start to be charged at time epoch t in region i. iii) Dispatch EVs to serve mobility demand $\boldsymbol{\xi}_t$. We denote the dispatching decision by $\boldsymbol{z}_t = \{z_t^{\ell,k} | \ell \in \mathcal{L}, k \in \mathcal{K}\}$, where $z_t^{\ell,k}$ represents the number of EVs in region i^ℓ with SoC k that are dispatched through link ℓ to serve demand ξ_t^ℓ. We denote the overall action of the system operator by $\boldsymbol{a}_t = \{\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t\} \in \mathcal{A}_t(\boldsymbol{s}_t)$, where $\mathcal{A}_t(\boldsymbol{s}_t)$ is the action space given state $\boldsymbol{s}_t \in \mathcal{S}_t$. Particularly, $\mathcal{A}_t(\boldsymbol{s}_t), t \in \mathcal{T}$, is characterized by the following constraints:

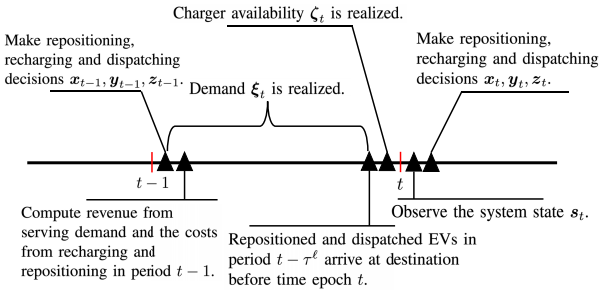


Fig. 2. An illustration of event sequence between time epochs $t-1$ and t. Right after time epoch $t-1$, the operator observes the system state in time period $t-1$ and makes operational decisions. Customers' arriving at any time between epoch $t-1$ and t contributes to the mobility demand in time period t. Right before time epoch t, some of the repositioned and dispatched EVs arrive at their destination region, and meanwhile, EVs under recharging finish their charging for one time period. Thus, the system state in time period t is realized, and the operator makes decisions in time period t.

$$\sum_{\ell \in \mathcal{L}^i} (x_t^{\ell,k} + z_t^{\ell,k}) + y_t^{i,k} \leq v_t^{i,k}, \quad \forall i \in \mathcal{M}, k \in \mathcal{K}, \tag{2}$$

$$z_t^{\ell,k} = 0, x_t^{\ell,k} = 0, \quad \forall k \in \mathcal{K} \setminus \mathcal{K}^\ell \tag{3}$$

$$\sum_{k \in \mathcal{K}} y_t^{i,k} \leq \zeta_t^i, y_t^{i,K} = 0, y_t^{i,k} \in \mathbb{N}, \quad \forall i \in \mathcal{M}, k \in \mathcal{K}, \tag{4}$$

$$\sum_{k \in \mathcal{K}^\ell} z_t^{\ell,k} \leq \xi_t^\ell, z_t^{\ell,k} \in \mathbb{N}, \quad \forall \ell \in \mathcal{L}, k \in \mathcal{K}^\ell. \tag{5}$$

Constraint (2) is the resource limit constraint, which shows that the total number of EVs to be dispatched, repositioned and recharged cannot exceed the number of available EVs $v_t^{i,k}$. In constraint (3), the dispatching decision $z_t^{\ell,k}$ and repositioning decision $x_t^{\ell,k}$ are forced to be zero when the SoC level k is insufficient to support the travel on link ℓ. Constraint (4) further restricts the number of EVs recharged in region i by the number of available chargers ζ_t^i. Constraint (5) ensures the total number of EVs dispatched through link ℓ is no more than the mobility demand ξ_t^ℓ. We depict the sequence of events in Fig. 2.

3) Transition Function: The state transition functions for the physical state from time epoch t to $t+1$ can be determined as follows:

$$v_{t+1}^{i,k} = v_t^{i,k} + (y_t^{i,k-1} - y_t^{i,k}) + \sum_{\ell \in \mathcal{L}_i} u_{t-\tau^\ell+1}^{\ell,k+k^\ell} - \sum_{\ell \in \mathcal{L}^i} (x_t^{\ell,k} + z_t^{\ell,k}), \quad \forall k \in \mathcal{K}, \tag{6}$$

$$u_t^{\ell,k} = x_t^{\ell,k} + z_t^{\ell,k}, \quad \forall \ell \in \mathcal{L}, k \in \mathcal{K}^\ell. \tag{7}$$

Equation (6) describes the dynamics of the number of EVs with SoC k in region i from time epoch t to $t+1$. Particularly, $y_t^{i,k-1}$ EVs are charged from SoC level $k-1$ to k, and $y_t^{i,k}$ EVs are charged from SoC level k to $k+1$ at time epoch $t+1$. In other words, the number of EVs with SoC k is increased by $y_t^{i,k-1}$, and meanwhile, decreased by $y_t^{i,k}$. Thus, the second term $y_t^{i,k-1} - y_t^{i,k}$ is the net increment of EVs in region i with SoC k due to recharging decisions. We denote the set

of links with a destination i by $\mathcal{L}_i$, such that $\sum_{\ell\in\mathcal{L}_i} u_{t-\tau^\ell+1}^{\ell,k+k^\ell}$ quantifies the number of in-transit EVs starting to travel at time epoch $t-\tau^\ell+1$ on link ℓ with an original SoC level $k+k^\ell$. By consuming k^ℓ SoC levels of energy and traveling time τ^ℓ, these EVs will arrive in region i at time epoch $t+1$ with a remaining SoC k. The last term of (6), $\sum_{\ell\in\mathcal{L}^i}(x_t^{\ell,k}+z_t^{\ell,k})$, is the number of EVs with SoC k leaving region i at time epoch t, due to repositioning and dispatching decisions. Equation (7) captures that the number of EVs starting to travel at time epoch t with SoC k on link ℓ equals the number of EVs with SoC k repositioned and dispatched on link ℓ.

4) Value Function: We formulate a finite horizon stochastic DP problem to maximize the total revenue of the EMoD system within T time periods. We first define the single-step system revenue $R_t(\boldsymbol{s}_t,\boldsymbol{a}_t)$ in time period t as,

$$R_t(\boldsymbol{s}_t,\boldsymbol{a}_t)=\sum_{\ell\in\mathcal{L}}\sum_{k\in\mathcal{K}^\ell} r_t^\ell z_t^{\ell,k}-\sum_{\ell\in\mathcal{L}}\eta_t^\ell\Big[\xi_t^\ell-\sum_{k\in\mathcal{K}^\ell} z_t^{\ell,k}\Big] -\sum_{\ell\in\mathcal{L}}\sum_{k\in\mathcal{K}^\ell}\gamma_t^\ell x_t^{\ell,k}-\sum_{i\in\mathcal{M}}\sum_{k\in\mathcal{K}}\lambda_t^i y_t^{i,k}, \tag{8}$$

which is the revenue earned by adopting action $\boldsymbol{a}_t$ given system state $\boldsymbol{s}_t$. Here r_t^ℓ and η_t^ℓ are the reward and penalty for serving or losing one customer on link ℓ in time period t. γ_t^ℓ is the cost for the per-EV repositioning on link ℓ, and λ_t^i denotes the per-EV recharging cost for one SoC level in region i. The single-step revenue (8) is computed by using the total revenue earned by serving the mobility demand through dispatching minus the penalty (i.e., the second term) for unserved demand, repositioning cost (i.e., the third term) and recharging cost (i.e., the fourth term).

Let $V_t(\boldsymbol{s}_t)$ denote the value function (or the profit-to-go function), which is the maximal expected revenue from time period t to the terminal period T. For any time period $t\in\mathcal{T}$, the revenue-maximizing problem of the EMoD system can be formulated as a finite horizon stochastic DP as follows:

$$V_t(\boldsymbol{s}_t)=\max_{\boldsymbol{a}_t\in\mathcal{A}_t(\boldsymbol{s}_t)} R_t(\boldsymbol{s}_t,\boldsymbol{a}_t)+\mathbb{E}[V_{t+1}(\boldsymbol{s}_{t+1})]. \tag{9}$$

The physical state $(\boldsymbol{v}_{t+1},\boldsymbol{u}_{1:t})$ at time epoch $t+1$ can be obtained from $(\boldsymbol{s}_t,\boldsymbol{a}_t)$ based on (6) and (7). The information state $(\boldsymbol{\xi}_{t+1},\boldsymbol{\zeta}_{t+1})$ is independent of the physical state $(\boldsymbol{v}_{t+1},\boldsymbol{u}_{1:t})$ and action $\boldsymbol{a}_{t+1}$, which can be regarded as external information. The second term on the right-hand side of (9) indicates the expected total revenue starting from period $t+1$ to T. With the probability of drawing $(\boldsymbol{\xi}_{t+1},\boldsymbol{\zeta}_{t+1})$ from $\mathcal{I}_{t+1}$ as $\mathbb{P}(\boldsymbol{\xi}_{t+1},\boldsymbol{\zeta}_{t+1})$, we can have $\mathbb{E}[V_{t+1}(\boldsymbol{s}_{t+1})]=\sum_{(\boldsymbol{\xi}_{t+1},\boldsymbol{\zeta}_{t+1})\in\mathcal{I}_{t+1}}\mathbb{P}(\boldsymbol{\xi}_{t+1},\boldsymbol{\zeta}_{t+1})V_{t+1}(\boldsymbol{s}_{t+1})$. However, solving (9) is highly nontrivial, due to the famous curse-of-dimensionality in DP [12]. The difficulties mainly come from two aspects: i) Given the system state $\boldsymbol{s}_t$, we need to solve an integer program to determine the optimal actions, which generally has high computational complexity, especially for large-scale systems. ii) Worse still, due to uncertain mobility demand and charger availability, problem (9) results in non-stationary decision policies, which requires us to repeatedly solve the computationally difficult problem for each possible state in each time period. Therefore, it is intractable to implement the standard backward induction method[2] in our problem. To tackle this challenging problem, we analyze the structural property of (9) and design an approximation algorithm to efficiently make operational decisions in the next section.

III. ADP-Based Solution Methodology

In this section, we propose an ADP-based algorithm to solve the DP problem (9). We first analyze the structural properties of this DP problem in Sec. III-A. Leveraging such properties, we design a piecewise linear function to approximate the value function in Sec. III-B. We then perform perturbation analysis to iteratively update the approximate function in Sec. III-C and present the complete algorithm for solving the DP problem of the EMoD system in Sec. III-D.

A. Structural Properties of Value Function

To approximate the value function $V_t(\boldsymbol{s}_t)$, we first demonstrate the structural properties of $V_t(\boldsymbol{s}_t)$ over the physical space $(\boldsymbol{v}_t,\boldsymbol{u}_{1:t-1})$ by the following Theorem 1. The information state $(\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ is captured as external information by our approximate value function in next subsection.

Theorem 1: The value function $V_t(\boldsymbol{s}_t),\forall t\in\mathcal{T}$ is coordinate-wise concave in $(\boldsymbol{v}_t,\boldsymbol{u}_{1:t-1})$, where a coordinate is referred to as one entry (e.g., $v_t^{i,k}$ or $u_t^{\ell,k}, i\in\mathcal{M},k\in\mathcal{K},\ell\in\mathcal{L},\iota\in[1,t-1]$) of the vector $(\boldsymbol{v}_t,\boldsymbol{u}_{1:t-1})$.

The proof of Theorem 1 makes use of the knowledge of convex optimization [24] and adopts an induction method. Note that we claim a discrete function is concave in one coordinate if its first-order difference is non-increasing in this coordinate.

Proof: To simplify the notation, we define the physical state as $\boldsymbol{p}_t:=(\boldsymbol{v}_t,\boldsymbol{u}_{1:t-1})$ and the system state as $\boldsymbol{s}_t:=(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$. Given $(\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$, we first analyze the continuous extension $\tilde{V}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ of the value function $V_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$. If we can prove that $\tilde{V}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ is concave in $\boldsymbol{p}_t$, then $V_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ is coordinate-wise concave in $\boldsymbol{p}_t$, since the first-order difference of $V_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ can be easily checked to be non-increasing in each coordinate of $\boldsymbol{p}_t$. To show the concavity of $\tilde{V}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$, we use induction:

(i) First we consider the base case for the terminal time period T. We define a continuous extension $\tilde{R}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t,\boldsymbol{a}_t)$ of the single-step revenue $R_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t,\boldsymbol{a}_t), t\in\mathcal{T}$. With $(\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t)$ as external parameters, the value of $\tilde{R}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t,\boldsymbol{a}_t)$ is determined by $(\boldsymbol{p}_t,\boldsymbol{a}_t)$, thus treated as a function of $(\boldsymbol{p}_t,\boldsymbol{a}_t)$. From (8), we can easily prove that $\tilde{R}_t(\boldsymbol{p}_t,\boldsymbol{\xi}_t,\boldsymbol{\zeta}_t,\boldsymbol{a}_t)$ is concave in $(\boldsymbol{p}_t,\boldsymbol{a}_t)$. By setting $\tilde{V}_{T+1}=0$ and $V_{T+1}=0$, we have

$$\tilde{V}_T(\boldsymbol{p}_T,\boldsymbol{\xi}_T,\boldsymbol{\zeta}_T)=\max_{a_T\in\tilde{\mathcal{A}}_T(\boldsymbol{p}_T)}\tilde{R}_T(\boldsymbol{p}_T,\boldsymbol{\xi}_T,\boldsymbol{\zeta}_T,\boldsymbol{a}_T). \tag{10}$$

Then the following Lemma 1 holds:

Lemma 1: $\tilde{V}_T(\boldsymbol{p}_T,\boldsymbol{\xi}_T,\boldsymbol{\zeta}_T)$ is concave in $\boldsymbol{p}_T$, where the feasible action space of $\boldsymbol{p}_T$, $\tilde{\mathcal{A}}_T(\boldsymbol{p}_T)$, is a polyhedron.

[2]The backward induction method is implemented by enumerating all possible outcomes from the current time to the end and then choosing the optimal path as the solution.

Proof: We have proved that $\tilde{R}_T$ is concave in $(\boldsymbol{p}_T, \boldsymbol{a}_T)$. With a convex action domain $\tilde{\mathcal{A}}_T(\boldsymbol{p}_t)$, $\tilde{V}_T(\boldsymbol{p}_T, \boldsymbol{\xi}_T, \boldsymbol{\zeta}_T)$ can be proved concave in $\boldsymbol{p}_T$ based on Chapter 3.2.5 in [24]. □

(ii) Next, we implement induction steps by assuming $\tilde{V}_{t+1}(\boldsymbol{p}_{t+1}, \boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})$ is concave over $\boldsymbol{p}_{t+1}$. We observe that the physical state $\boldsymbol{p}_{t+1}$ can be obtained from an affine mapping $\boldsymbol{p}_{t+1} = f(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t)$ based on (6) and (7). $\tilde{V}_{t+1}(f(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t), \boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})$ can be proved concave in $\boldsymbol{p}_t$, since the second-order derivatives of $f(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t)$ and $\tilde{V}_{t+1}(f(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t), \boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})$ to $\boldsymbol{p}_t$ are zero and negative, respectively. We define a Q-function of a state-action pair as

$$\begin{aligned} Q_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t) = \; & \tilde{R}_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t) \\ & + \mathbb{E}[\tilde{V}_{t+1}(\boldsymbol{p}_{t+1}, \boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})], \\ & \forall \boldsymbol{a}_t \in \tilde{\mathcal{A}}_t(\boldsymbol{p}_t), \quad \forall \boldsymbol{s}_t \in \mathcal{S}. \end{aligned}$$

Since taking expectation over $(\boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})$ preserves the concavity of $\tilde{V}_{t+1}(\boldsymbol{p}_{t+1}, \boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1})$ in $\boldsymbol{p}_t$, $Q_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t)$ can be proved concave in $(\boldsymbol{p}_t, \boldsymbol{a}_t)$. Following the same argument in the proof of Lemma 1, we can show $\tilde{V}_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ is concave in $\boldsymbol{p}_t$,

$$\tilde{V}_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t) = \max_{\boldsymbol{a}_t \in \tilde{\mathcal{A}}_t(\boldsymbol{p}_t)} Q_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t, \boldsymbol{a}_t). \tag{11}$$

Consequently, the integral value function $V_t(\boldsymbol{p}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ can be proved coordinate-wise concave in $\boldsymbol{p}_t$, which concludes the proof of Theorem 1. □

B. Separable PieceWise Linear Approximation

Based on the coordinate-wise concavity shown in Theorem 1 and the fact that our DP formulation has integer solutions, we approximate the value function (9) by a separable piecewise linear function.

In previous discussion, we only prove that the value function is coordinate-wise concave in the physical state. However, we also want to incorporate the information state in our approximate value function. By randomly sampling the information state $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ from $\mathcal{I}_t$, we can categorize different values of $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ into different information intervals. We denote the index set of information interval by $\mathcal{C} := \{1, \ldots, C\}$. The number of intervals C can be easily tuned by adjusting different ranges of demand and number of chargers. After that, we can design the approximate value function on different information intervals to capture both physical state and information state.

For notational convenience, we present the physical state $\boldsymbol{p}_t$ by listing each of its coordinates. In this way, we have $\boldsymbol{p}_t := (p_t^1, \ldots, p_t^D)$, where D is the dimension of $\boldsymbol{p}_t$. To coincide with our definition of $\boldsymbol{p}_t := (\boldsymbol{v}_t, \boldsymbol{u}_{1:t-1})$ in Sec. II-A, the dimension is calculated by $D = M \times (K+1) + L \times (K+1) \times (t-1)$. We denote the coordinate index set by $\mathcal{D} := \{1, \ldots, D\}$, and $p_t^d, d \in \mathcal{D}$, is an integer value referring to the number of EVs shown in the d-th coordinate of $\boldsymbol{p}_t$. For example, the 1-st coordinate of $\boldsymbol{p}_t$ (i.e., p_t^1) represents the value of $v_t^{1,0}$, which is the number of EVs in region 1 with SoC level 0.

For each information interval $c \in \mathcal{C}$ and coordinate of the value function $d \in \mathcal{D}$, we define a piecewise linear function $g_{d,t}^c(p_t^d)$ to approximate the system benefit of holding p_t^d EVs. Specifically, the value of $g_{d,t}^c(p_t^d)$ is determined by a vector of slopes $\boldsymbol{\theta}_{d,t}^c := \{\theta_{d,t}^c(j) | j \in \mathcal{J} := \{1, 2 \ldots, N\}\}$, where $\theta_{d,t}^c(j)$ is the slope of the j-th segment of $g_{d,t}^c(\cdot)$. To maintain the coordinate-wise concavity of the value function, $\theta_{d,t}^c(j)$ should be non-increasing over $j \in \mathcal{J}$.

If we set the horizontal segment length of $g_{d,t}^c(\cdot)$ as 1, $\boldsymbol{\theta}_{d,t}^c$ can be represented by $\boldsymbol{\theta}_{d,t}^c := (\theta_{d,t}^c(1), \theta_{d,t}^c(2), \ldots, \theta_{d,t}^c(N))$, and we have

$$g_{d,t}^c(p_t^d) = \sum_{j=1}^{p_t^d} \theta_{d,t}^c(j), \tag{12}$$

which is the sum of slopes up to the value p_t^d. For a total of D piecewise linear functions of state $\boldsymbol{p}_t$, the approximate value function is given by

$$\hat{V}_t(\boldsymbol{p}_t, \boldsymbol{\theta}_t^c) = \sum_{d=1}^{D} g_{d,t}^c(p_t^d) = \sum_{d=1}^{D} \sum_{j=1}^{p_t^d} \theta_{d,t}^c(j), \tag{13}$$

where $\boldsymbol{\theta}_t^c := \{\boldsymbol{\theta}_{d,t}^c | d \in \mathcal{D}\}$, and the information state $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ falls in the c-th interval. Note that the value function can also be approximated as other forms of piecewise linear functions by setting the segment length of $g_{d,t}^c(\cdot)$ to different values.

Next, we proceed to show how to iteratively update the slope $\boldsymbol{\theta}_t^c$ in $\hat{V}_t(\boldsymbol{p}_t, \boldsymbol{\theta}_t^c)$ based on perturbation analysis.

C. Perturbation Analysis

Using (13) to approximate the value function, we need to estimate a large number of parameters, which is highly non-trivial.

Recall that the parameters in (13) are the slope $\boldsymbol{\theta}_t^c$ of a piecewise linear function, which indicates the incremental value of having one additional EV. Thus, we propose a value function approximation algorithm based on perturbation analysis [25], [26] and update $\boldsymbol{\theta}_t^c$ iteratively.

In each iteration, we first replace the value function $V_{t+1}(\boldsymbol{s}_{t+1})$ in (9) by $\hat{V}_{t+1}(\boldsymbol{p}_{t+1}, \boldsymbol{\theta}_{t+1}^b)$, which is our current approximation of $V_{t+1}(\boldsymbol{s}_{t+1})$ obtained in the last iteration. Here b is the information index in time period $t+1$. Specifically, we solve (14) to get the objective value $\bar{V}_t(\boldsymbol{s}_t)$ as an estimate of $V_t(\boldsymbol{s}_t)$,

$$\begin{aligned} \bar{V}_t(\boldsymbol{s}_t) &= \max_{\boldsymbol{a}_t \in \mathcal{A}_t(\boldsymbol{s}_t)} R_t(\boldsymbol{s}_t, \boldsymbol{a}_t) + \mathbb{E}[\hat{V}_{t+1}(\boldsymbol{s}_{t+1}, \boldsymbol{\theta}_{t+1}^b)] \\ &= \max_{\boldsymbol{a}_t \in \mathcal{A}_t(\boldsymbol{s}_t)} R_t(\boldsymbol{s}_t, \boldsymbol{a}_t) + \frac{1}{H} \sum_{h=1}^{H} \sum_{d=1}^{D} \sum_{j=1}^{p_{t+1}^d} \theta_{d,t+1}^{b_h}(j), \end{aligned} \tag{14}$$

where the expectation is calculated by randomly sampling $(\boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1}) \in \mathcal{I}_{t+1}$ for H times, and b_h is the information interval index obtained in the h-th sampling. The hidden information for state $\boldsymbol{s}_t$ is the information index c.

In order to solve (14), we equivalently transform (14) into the following LP problem:

$$\max_{\boldsymbol{a}_t \in \mathcal{A}_t, q_j^d} \; R_t(\boldsymbol{s}_t, \boldsymbol{a}_t) + \frac{1}{H} \sum_{h=1}^{H} \sum_{d=1}^{D} \sum_{j=1}^{N} \theta_{d,t+1}^{b_h}(j) \cdot q_j^d, \tag{15a}$$

$$\text{s.t. } q_j^d \in \{0,1\}, \quad \forall d \in \mathcal{D}, j \in \mathcal{J}, \tag{15b}$$

$$\sum_{j=1}^{N} q_j^d = p_{t+1}^d, \quad \forall d \in \mathcal{D}, \tag{15c}$$

$$constraints : (6), (7)$$

where q_j^d is the newly introduced binary variable and indicates if the j-th slope should be counted into the objective (15a). In constraint (15c), given $(\boldsymbol{s}_t, \boldsymbol{a}_t)$, p_{t+1}^d is derived from the state transition equations of (6) and (7). Note that q_j^d solely depends on the physical state $p_{t+1}^d, d \in \mathcal{D}$, and thus q_j^d is the same for any information interval index b, as long as $(\boldsymbol{s}_t, \boldsymbol{a}_t)$ is given.

To show the equivalence between (14) and (15), we present the following Lemma 2.

Lemma 2: For any state $\boldsymbol{s}_t$, the optimal solution q_j^{d*} of (15) *is*

$$q_j^{d*} = \begin{cases} 1, & j \in \{1, \dots, p_{t+1}^d\}, \\ 0, & otherwise, \end{cases}$$

such that $\sum_{j=1}^{N} \theta_{d,t+1}^{b_h}(j) \cdot q_j^{d*} = \sum_{j=1}^{p_{t+1}^d} \theta_{d,t+1}^{b_h}(j)$.

Proof: Suppose the optimal solution has one $q_{j_1}^{d*} = 1$ for $j_1 \in \mathcal{J} \setminus \{1, \dots, p_{t+1}^d\}$. To satisfy constraint (15c), there is at least one $q_{j_2}^{d*} = 0$ for $j_2 \in \{1, \dots, p_{t+1}^d\}$. Given $(\boldsymbol{s}_t, \boldsymbol{a}_t)$, we can always improve the objective (15a) by setting $q_{j_1}^{d*} = 0$ and $q_{j_2}^{d*} = 1$ since $\theta_{d,t+1}^{b_h}(j)$ is non-increasing over $j \in \mathcal{J}$, which contradicts with the optimality of $q_j^{d*}, j \in \mathcal{J}$. □

Note that the above transformation also works for the general piecewise linear approximation function with different segment lengths $\rho(j) \in [0, N], j \in \mathcal{J}, \sum_{j \in \mathcal{J}} \rho(j) = N$. In that case, we have $q_j^d \in \{0, 1, \dots, \rho(j)\}$ and q_j^{d*} is the optimal value of p_{t+1}^d distributed in the j-th segment with $\sum_{j \in \mathcal{J}} q_j^{d*} = p_{t+1}^d$.

Given the system state $\boldsymbol{s}_t$, we can obtain the optimal action $\boldsymbol{a}_t^*$ for (14) by solving (15) with commercial solvers. After obtaining the objective value $\bar{V}_t(\boldsymbol{s}_t)$ of (14), we perturb the physical state $\boldsymbol{p}_t$ by $\boldsymbol{p}_{t,d} = \boldsymbol{p}_t + \boldsymbol{e}_d$ to calculate the marginal value at point p_t^d on coordinate $d, \forall d \in \mathcal{D}$. Here $\boldsymbol{e}_d$ is referred to as a D-dimensional column vector, with only the d-th element equal to 1 and all others equal to zero. In other words, $\boldsymbol{p}_{t,d}$ equals $\boldsymbol{p}_t$, except that the d-th coordinate of $\boldsymbol{p}_t$ is incremented by 1. The system state after perturbation is denoted by $\boldsymbol{s}_{t,d} = (\boldsymbol{p}_{t,d}, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$. Note that, by having one additional EV, the perturbation may result in an infeasible state $\boldsymbol{s}_{t,d}$, which violates the constraint of the total number of EVs in (1). However, this state does not exist in real operation process, which is only used to estimate the marginal value of state $\boldsymbol{s}_t$.

With the information state $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$ and its index c unchanged by perturbation, we calculate the value of perturbation by $\Delta_{d,t}^c(\boldsymbol{s}_t) = \bar{V}_t(\boldsymbol{s}_{t,d}) - \bar{V}_t(\boldsymbol{s}_t), \forall d \in \mathcal{D}$. The corresponding slope is then iteratively updated by

$$\theta_{d,t}^c(p_t^d) = (1 - \alpha^w)\theta_{d,t}^c(p_t^d) + \alpha^w \Delta_{d,t}^c(\boldsymbol{s}_t), \tag{16}$$

where α is a constant less than one. w is the iteration index, indicating the total number of rounds that have been implemented during the updating process. α^w is called the step size, which decreases exponentially in the rounds of iterations.

Algorithm 1 EMoD Value Function Approximation Algorithm

1: Initialize $\theta_{d,t}^c(j), \forall t \in \mathcal{T}, c \in \mathcal{C}, d \in \mathcal{D}, j \in \mathcal{J}$.
2: **for** $w = 1, 2, \dots, W$ **do**
3: Set step size α^w.
4: **for** $t = 1, 2, \dots, T$ **do**
5: Randomly sample $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t) \in \mathcal{I}_t$ and get $c \in \mathcal{C}$.
6: **if** $t = 1$ **then**
7: Randomly generate an initial physical state $\boldsymbol{p}_t^w$.
8: **else**
9: Compute $\boldsymbol{p}_t^w = f(\boldsymbol{s}_{t-1}^w, \boldsymbol{a}_{t-1}^{w,*})$.
10: **end if**
11: Randomly sample $(\boldsymbol{\xi}_{t+1}, \boldsymbol{\zeta}_{t+1}) \in \mathcal{I}_{t+1}$ for H times.
12: For $\boldsymbol{a}_t^w \in \mathcal{A}_t(\boldsymbol{s}_t^w)$, calculate $\bar{V}_t(\boldsymbol{s}_t^w)$ and obtain $\boldsymbol{a}_t^{w,*}$ by solving (14).
13: **for** $d = 1, 2, \dots, D$ **do**
14: Set $\boldsymbol{s}_t = \boldsymbol{s}_t^w$.
15: Randomly perturb the state $\boldsymbol{s}_t$ by $\boldsymbol{s}_{t,d}$.
16: Compute $\Delta_{d,t}^c(\boldsymbol{s}_t) = \bar{V}_t(\boldsymbol{s}_{t,d}) - \bar{V}_t(\boldsymbol{s}_t)$.
17: Set $\theta_{d,t}^c(p_t^d) = (1 - \alpha^w)\theta_{d,t}^c(p_t^d) + \alpha^w \Delta_{d,t}^c(\boldsymbol{s}_t)$.
18: Execute SPAR algorithm on $\boldsymbol{\theta}_{d,t}^c$.
19: **end for**
20: **end for**
21: **end for**

Note that the update (16) cannot guarantee $\theta_{d,t}^c(j)$ is non-increasing over $j \in \mathcal{J}$, which may lead to a non-concave function $g_{d,t}^c(\cdot)$. Therefore, we apply the SPAR algorithm [12] for updating slopes $\boldsymbol{\theta}_{d,t}^c$ to maintain the concavity of $g_{d,t}^c(\cdot)$. Here the SPAR algorithm is used to check if the slopes $\theta_{d,t}^c(j)$ are non-increasing over $j \in \mathcal{J}$. If some slopes $\theta_{d,t}^c(j)$ are increasing over $j \in \{n_1, n_1 + 1 \dots, n_2\} \subseteq \mathcal{J}$, the SPAR calculates the average of these slopes and set $\theta_{d,t}^c(j), j \in \{n_1, \dots, n_2\}$ as the average value. The SPAR keeps checking $\theta_{d,t}^c(j)$ from $j = 1$ to N until the non-increasing property is retrieved for all slopes.

D. ADP-Based Solution

In this section, we show the complete solution for the EMoD problem, which consists of two main steps.

1) Value Function Approximation: We approximate the value function (9) by a piecewise linear function (13) using Algorithm 1. We uniformly initialize all slopes $\theta_{d,t}^c(j)$ for $\forall t \in \mathcal{T}, c \in \mathcal{C}, d \in \mathcal{D}, j \in \mathcal{J}$ and then iteratively update the slopes in a total of W iterations. In each iteration w, the step size is set as α^w. We randomly generate the physical state $\boldsymbol{p}_t^w$ if we are at the initial period 1, otherwise, calculate it by $\boldsymbol{p}_t^w = f(\boldsymbol{s}_{t-1}^w, \boldsymbol{a}_{t-1}^{w,*})$. Here $\boldsymbol{a}_{t-1}^{w,*}$ is the optimal action derived by solving (14). The information state $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t), \forall t \in \mathcal{T}$ is randomly sampled from $\mathcal{I}_t$ and the corresponding information index c can be observed. In Line 12, we calculate $\bar{V}_t(\boldsymbol{s}_t^w)$ and obtain the optimal action $\boldsymbol{a}_t^{w,*}$ of (14) by utilizing the slopes $\boldsymbol{\theta}_{t+1}^b$ updated in the last iteration. After that, we perturb the state $\boldsymbol{s}_t^w$ in each coordinate $d \in \mathcal{D}$ of $\boldsymbol{p}_t$ in Line 15 and update the slopes $\theta_{d,t}^c(p_t^d)$ in Lines 16 to 17. We apply the

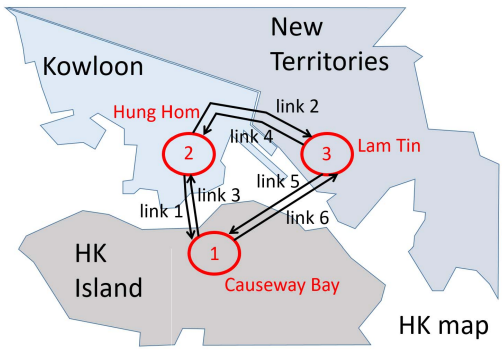


Fig. 3. A HK map with three regions and six links.

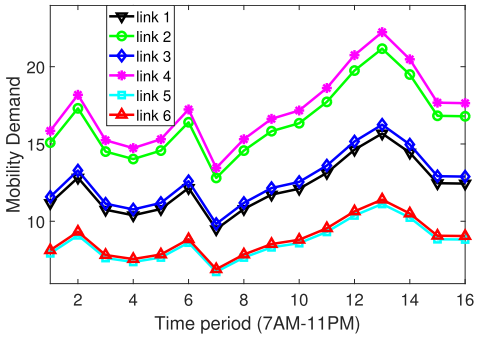
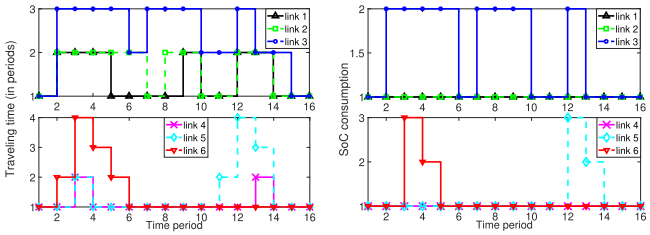


Fig. 4. Mobility demand in each time period.



(a) Traveling time (b) SoC consumption

Fig. 5. Time-varying traveling time and SoC consumption.

SPAR algorithm to maintain the coordinate-wise concavity of $\hat{V}_t(\boldsymbol{p}_t, \boldsymbol{\theta}_t^c)$ in Line 18, and move on to the next period until the end period T. Note that $\theta_{d,T+1}^c(j) = 0$, $j \in \mathcal{J}$. This procedure is repeated until finishing W rounds of updating.

This step is implemented in an offline manner. The information state $(\boldsymbol{\xi}_t, \boldsymbol{\zeta}_t), \forall t \in \mathcal{T}$ is sampled from the distributions of mobility demand and charger availability estimated from historical data. Since such distributions will usually not change dramatically on an ordinary day, the offline value function approximation can be implemented with a relatively low frequency (e.g., weekly or monthly). In special cases (e.g., public holidays or sports events), Algorithm 1 can be run based on the specific data from these cases.

2) Real-Time Decision-Making: By implementing Algorithm 1 in an offline manner, we can obtain the approximate value function $\hat{V}_t(\boldsymbol{p}_t, \boldsymbol{\theta}_t^c), \forall t \in \mathcal{T}, c \in \mathcal{C}$. Then we substitute the approximate value function to the DP and can directly solve the problem (14) (which is equivalent to the LP (15)) to obtain the real-time decisions within seconds. In each step, the expected value function $\mathbb{E}[\hat{V}_{t+1}(\boldsymbol{s}_{t+1}, \boldsymbol{\theta}_{t+1}^b)]$ is still estimated by the sampled average value while the state transitions across steps are determined by the actual realizations of the mobility demand and charger availability.

IV. Simulation Results

In this section, we assess the performance of our algorithm in an EMoD system in three service regions, Causeway Bay, Hung Hom and Lam Tin, of Hong Kong (HK) in Fig. 3.

A. Simulation Setups

EMoD System Model. We consider EMoD systems with two different scales:

1) A small-scale system operates a total of $N = 5$ EVs and provides on-demand mobility services for $L = 2$ links in $M = 2$ service regions (Hung Hom and Lam Tin). The total number of SoC levels is set as $K = 3$ and the service time horizon is $T = 3$.
2) A large-scale system operates a total of $N = 100$ EVs and provides on-demand mobility services for $L = 6$ links in $M = 3$ service regions (Causeway Bay, Hung Hom and Lam Tin). The total number of SoC levels is set as $K = 4$ and the service time horizon is $T = 16$.

The small-scale system is used to test the accuracy of our ADP-based solution, by comparing it with the optimal solution from backward induction. The large-scale system is to test the effectiveness of our algorithm in making operational decisions in practice by comparing the overall system revenues obtained by our algorithm and two benchmark algorithms.

System State. We introduce detailed large-scale system settings and omit those of the small-scale system, which are obtained with the same logic in the large-scale system. We obtain the physical state by randomly generating the number of EVs with different SoC levels in different locations. To estimate the mobility demand on different travel links, we use the data from HK Transport Department [27], [28]. Specifically, based on the proportion of total traveling of one day in each hour (7AM-11PM), we estimate the mobility demand distribution. The time duration of one time period is set as one hour. By further accessing the average daily traffic on the major roads between Causeway Bay, Hung Hom and Lam Tin, we can obtain the hourly mean demand on each travel link. To fit the number of 100 EVs, we scale the total demand of the six links in the peak hour 8-9 AM (period 2) as 80% of the total EVs and present the obtained demand patterns in Fig. 4. We also add a margin of 20% of the mean to capture the demand variance, and sample the demand uniformly from the corresponding margin.

We also use the data from HK Environmental Protection Department for the total number of public EV chargers in the three areas (HK Island, New Territories and Kowloon) of HK [29]. The total number of public EV chargers in the three areas are 634, 861, and 1473, respectively, and the total number of registered EVs in HK is around 7000. Thus, by introducing a scaling factor of 70, the total number of

TABLE II
DISPATCHING REWARD AND REPOSITIONING COST (HKD)

	High Congestion	Medium Congestion	Low Congestion
r_t^ℓ	$\{54, 30, 18, 17, 32, 50\}$	$\{36, 36, 16, 16, 20, 35\}$	$\{18, 17, 16, 15, 17, 17\}$
γ_t^ℓ	$\{24, 16, 8, 8, 14, 23\}$	$\{16, 12, 7, 7, 10, 15\}$	$\{12, 10, 7, 27, 10, 11\}$

public chargers in the three service regions of the EMoD system are estimated as $\{10, 13, 22\}$. The distributions of available chargers in each period are differentiated by high-demand periods $t \in \mathcal{T}_h := \{1, 2, 3, 12, 13, 14\}$ and low-demand periods $t \in \mathcal{T} \setminus \mathcal{T}_h$. Specifically, the number of available chargers in the three service regions is drawn from $[0, 7]$, $[0, 9]$ and $[0, 15]$ in the high-demand periods and $[4, 10]$, $[6, 13]$ and $[10, 22]$ in the low-demand periods, since it's possible to have no available chargers in high-demand periods and abundant available chargers in low-demand periods.

Traffic and Price. The traveling times between Causeway Bay, Hung Hom and Lam Tin are obtained from Google Map data (7AM-11PM). By introducing a scaling factor of 12, 5-minute traveling time is scaled to one hour (one time period), in order to capture the dynamic of in-transit EVs between different time periods. Note that we can also set a shorter time duration of one time period to fit the actual traveling time. Energy consumption is set proportional to the traveling time. By scaling and ceiling the data to integer values, the traveling time (number of periods) and energy consumption (SoC levels) are visualized in Fig. 5.

The EVs have a battery capacity of 40 kWh, and the charging power is set as 10 kW. Based on the electricity prices in the three areas of HK, the per-SoC (10 kWh) charging costs are set as $\{\lambda_t^i\}_{i\in\mathcal{M}} = \{10, 9, 8\}, \forall t \in \mathcal{T}$ (HKD). We also set the penalty of losing one customer as $\{\eta_t^\ell\}_{\ell\in\mathcal{L}} = \{5, 6, 6, 5.5, 6, 5\}, \forall t \in \mathcal{T}$ (HKD). We set the per-EV dispatching reward r_t^ℓ and per-EV repositioning cost γ_t^ℓ on each travel link $\ell \in \mathcal{L}$ based on traffic congestion conditions, which are distinguished by high-congestion $t \in \{2, 3, 12, 13\}$, medium-congestion $t \in \{4, 5, 11, 14\}$ and low-congestion periods $t \in \{1, 6, 7, 8, 9, 10, 15, 16\}$. We list the setup information of r_t^ℓ and γ_t^ℓ in Table II.

B. Benchmarks

We compare our proposed algorithm with the following two model-free benchmark algorithms.

1) Q-Learning With Linear Function Approximation: This algorithm approximates the Q-function $Q_t(\boldsymbol{s}_t, \boldsymbol{a}_t)$ by a linear feature based approximate function $\tilde{Q}_t(\boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{\omega}_t) = \sum_{\beta=1}^{\bar{\beta}} \omega_t^\beta \phi_t^\beta(\boldsymbol{s}_t, \boldsymbol{a}_t) = \boldsymbol{\omega}_t^\intercal \boldsymbol{\Phi}_t(\boldsymbol{s}_t, \boldsymbol{a}_t)$, where $\boldsymbol{\Phi}_t(\boldsymbol{s}_t, \boldsymbol{a}_t)$ is the feature function with $\bar{\beta} = 24$ dimensions, representing the mapping from the state and action space to the feature space, and $\boldsymbol{\omega}_t$ is the weight vector for the feature. The 24 features $\phi_t^\beta(\boldsymbol{s}_t, \boldsymbol{a}_t)$ are extracted as follows: $\sum_{k\in\mathcal{K}^\ell} x_t^{\ell k}$, $\sum_{k\in\mathcal{K}^\ell} z_t^{\ell k}$ and $\sum_{\iota=t-\tau^\ell+1}^{t-1} \sum_{k\in\mathcal{K}^\ell} u_\iota^{\ell k}, \forall \ell \in \mathcal{L}$ are the numbers of repositioned, dispatched and in-transit EVs on each link ℓ. $\sum_{k\in\mathcal{K}} y_t^{ik}, \forall i \in \mathcal{M}$ captures the number of EVs that start to be charged in period t, and $\sum_{k\in\mathcal{K}} v_t^{ik}, \forall i \in \mathcal{M}$ indicates the total number of EVs in each region. The weight vector $\boldsymbol{\omega}_t$ can then be obtained from a temporal difference learning process. Interested readers can refer to [30] for more about the aforementioned method.

2) Greedy Heuristic Algorithm: This algorithm serves the mobility demand by dispatching EVs with sufficient SoC levels from low to high, so as to reserve EVs with high SoC levels for future usages. A repositioning decision is made after dispatching. Specifically, remaining EVs with sufficient SoCs will be repositioned to the adjacent region with the highest demand in last period. Without exact information of the number of chargers in the future, the remaining EVs after repositioning will be recharged based on current SoC levels. EVs with low SoCs have high priority to be charged by the available chargers.

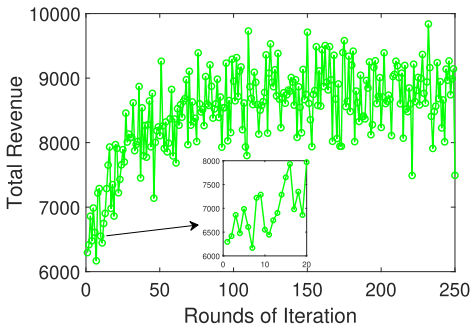


Fig. 6. Performance improvement during slope updating.

C. Parameter Update

To evaluate the performance of Algorithm 1, we first present the revenue improvement during parameter updating iterations in the large-scale system. The slopes $\theta_{d,t}^c(j), \forall t \in \mathcal{T}, d \in \mathcal{D}, c \in \mathcal{C}, j \in \mathcal{J}$ are initialized as 1, and α is set as 0.99.

Fig. 6 shows the overall system revenue within 16 periods versus the rounds of iterations for the proposed algorithm. We can observe that during 250 rounds of updating, the total revenue fluctuates, since the revenue highly depends on the initial states and the realization of aforementioned uncertain information. However, with the increase of the iterations, the revenue increases rapidly and then fluctuates around a constant value. The running time for 250 iterations is around 14 hours. However, this algorithm is executed in an offline manner (e.g., day-ahead or week-ahead), where we can obtain the approximate value function (14). Based on the approximate value function, we can make real-time decisions within seconds by solving the LP (15).

D. Accuracy of the Approximation

We present the accuracy of our approximate function in the tests of the small-scale system. Specifically, we compare the overall system revenues of Algorithm 1, Q-learning, Greedy Heuristic with OPT, the revenue obtained by the optimal backward induction [11]. Note that even for a small-scale system, the running time of the backward induction for each initial state is more than 5 hours, since we have about 10^6 possible actions.

For any algorithm, the revenue is calculated based on a given realization of the uncertainties. For the backward

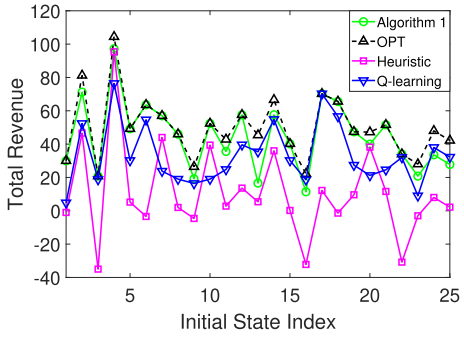


Fig. 7. Comparison of revenues in the small-scale system.

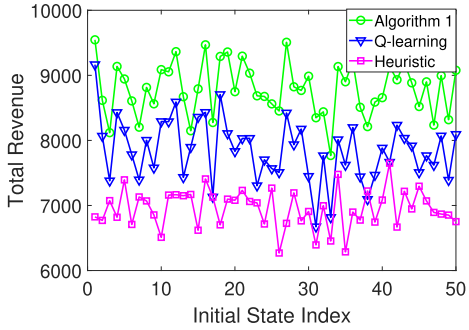


Fig. 8. Comparison of revenues between Algorithm 1 and two benchmark algorithms in the large-scale system.

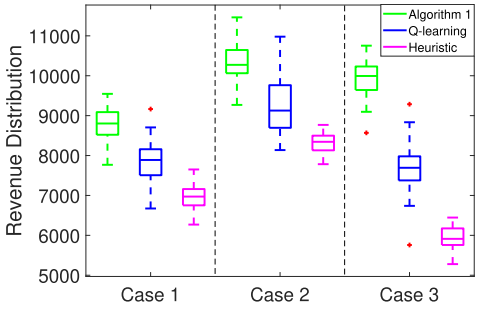
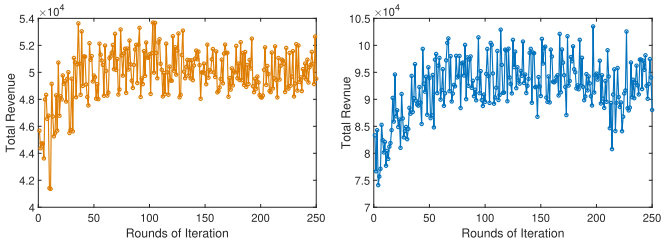


Fig. 9. Revenue distributions under different demand patterns in the large-scale system.



(a) The test on 500 EVs (b) The test on 1000 EVs

Fig. 10. Tests on different numbers of EVs.

induction, we can get the maximal total revenue based on the knowledge of the whole uncertainty realization at the beginning of the day. On the contrary, when running Algorithm 1, Q-learning and Greedy Heuristic, the future realization is not known. Thus, by enumerating all possible decisions and outcomes, the revenue of the backward induction provides the optimal benchmark for a given realization.

Fig. 7 shows the revenue comparison for different initial states. Specifically, we randomly generate an initial state and compute the expected revenue of 5 realizations of future uncertainties. We can observe that Algorithm 1 yields near-optimal solution compared with the optimal solution obtained by backward induction and outperforms the other two benchmarks. This result demonstrates the accuracy of our approximation in Algorithm 1. Since we have a penalty for unserved demand in our formulation, it's possible to have negative overall revenue in the small-scale system.

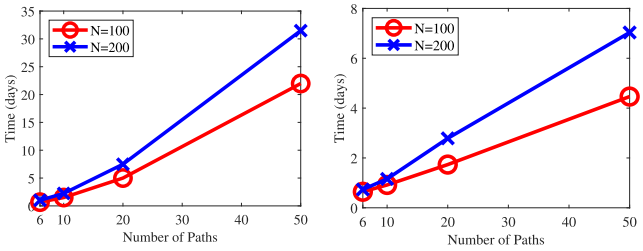
E. Effectiveness of the Approximation

Next, we show the effectiveness of our algorithm in making decisions in the large-scale system. Fig. 8 compares the revenues of Algorithm 1, Q-learning and Greedy Heuristic algorithm for 50 randomly generated initial states. For each initial state, we calculate the expected revenue for 5 future realizations. We can find our algorithm produces better revenue than the other two algorithms.

Based on our formulation (8), we notice that the repositioning cost and demand-supply mismatch penalty can greatly affect the overall revenue, and these costs highly depend on the temporal and spatial difference of mobility demand. Thus, to evaluate the impact of different demand patterns, we intentionally tune the means of demand in different regions in the following cases. In case 1 (case 3), we set larger (smaller) mean demands on travel links originating from region 2 than those in the other two regions (the means are the same in these two regions). In case 2, we set the same mean in all links. In each case, the total (mean) demand of all links in time period 2 is set as 80% of the total number of EVs. The mean demand in each time period is then set based on the demand distribution obtained in Sec. IV-A. The distributions of available chargers in each period are the same as our previous settings in Sec. IV-A. Fig. 10 shows the revenue distributions of the three algorithms. In each case, we generate different initial states and compute the expected revenue for uncertainty realizations under different demand patterns. We can observe i) the revenue distribution under the balanced demand is much better than those in the other two cases, since the repositioning cost and penalty are decreased, and ii) Algorithm 1 significantly outperforms the other algorithms in all cases.

F. Scalability of Algorithm 1

In order to show the scalability of Algorithm 1, we first perform tests on systems with 500 and 1000 EVs, respectively. The hourly mean demand and number of available chargers are scaled by 5 and 10, according to the number of EVs. We show the overall system revenues during the offline updating processes in Fig. 10. We can observe that the revenues increase and then fluctuate around constant values in both cases. The execution times of the two tests are 26 and 42 hours, respectively. Further, we show the execution time of Algorithm 1 (250 iterations) as the number of paths



(a) Algorithm 1 (b) Algorithm 1 with $\hat{D}$ update

Fig. 11. Execution time of approximate value functions as the number of paths increases.

varies from 6 to 50 in Fig. 11(a). We scale the transportation network by adding one service region with an increase of two paths. It can be observed that the execution time increases up to 25 days when the system is with 25 regions, 50 paths and 200 EVs. However, this is acceptable since Algorithm 1 is implemented in an offline manner (on a weekly or monthly basis) to obtain the slope values of the approximate value function.

There exist many approaches to further improve the scalability in practice. We note that the most time-consuming step of Algorithm 1 lies in the slope updates of D coordinates (Lines 13-19). We can randomly choose $\hat{D}$ coordinates from all the D coordinates to update the slopes and keep other slopes unchanged in each iteration. In this way, the execution time can be significantly reduced to a few days as shown in Fig. 11(b), where we set $\hat{D} = 10$. However, this solution-wise heuristics will sacrifice the accuracy of the approximate value function and the resulting system revenues (around 73% of the revenue obtained by running Algorithm 1). The approximation accuracy is expected to further reduce in larger systems since $\hat{D}$ updates will then provide much coarse slope updates to approximate the value functions. In addition, the execution time can be reduced by employing more computation resources, when the operator is faced with a larger system. We can also partition a large geographic area into several small areas and provide mobility services separately for each area to speed up the execution of Algorithm 1.

V. Conclusion

In this paper, we proposed a mathematical model for mobility and energy management in an EMoD system under stochastic and non-stationary mobility demand and charger availability. Specifically, the management of the EMoD system involves a complex sequential decision-making process, including repositioning, recharging and dispatching. Thus, we formulated the problem into a stochastic DP problem. To tackle the curse-of-dimensionality in solving the large-scale DP problem, we adopted a separable piecewise linear function as an approximation of the value function, based on a rigorously proved structural property of the formulated problem. Moreover, we designed an algorithm to update the parameters of the approximate value function iteratively in an offline manner. During the operation of the EMoD system, we can apply this approximate value function to make decisions within seconds. In numerical tests, we conducted both small-scale and large-scale tests to show the accuracy and effectiveness of our algorithm by comparing overall system revenues with different benchmarks.

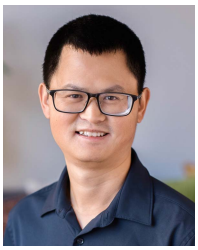
References

[1] H. Ritchie and M. Roser, "Urbanization," in *Our World in Data*. 2018.
[2] G. Martine and A. Marshall, "State of world population 2007: Unleashing the potential of urban growth, report," United Nations Population Fund, New York, NY, USA, Tech. Rep., 2007.
[3] W. J. Mitchell, C. E. Borroni-Bird, and L. D. Burns, *Reinventing the Automobile: Personal Urban Mobility for the 21st Century*. Cambridge, MA, USA: MIT Press, 2010.
[4] S. Shaheen, A. Cohen, B. Yelchuru, S. Sarkhili, and B. A. Hamilton, "Mobility on demand operational concept report," United States Dept. Transp. Intell. Transp., Tech. Rep., 2017.
[5] J. N. Bajpai, "Emerging vehicle technologies & the search for urban mobility solutions," *Urban, Planning Transp. Res.*, vol. 4, no. 1, pp. 83–100, Jan. 2016.
[6] S. Jiang, L. Chen, A. Mislove, and C. Wilson, "On ridesharing competition and accessibility: Evidence from Uber, Lyft, and taxi," in *Proc. World Wide Web Conf.*, 2018, pp. 863–872.
[7] X. Tan and A. Leon-Garcia, "Autonomous mobility and energy service management in future smart cities: An overview," in *Proc. 4th Int. Conf. Universal Village (UV)*, Oct. 2018, pp. 1–6.
[8] R. Nair and E. Miller-Hooks, "Fleet management for vehicle sharing operations," *Transp. Sci.*, vol. 45, no. 4, pp. 524–540, 2011.
[9] R. Zhang and M. Pavone, "Control of robotic mobility-on-demand systems: A queueing-theoretical perspective," *Int. J. Robot. Res.*, vol. 35, nos. 1–3, pp. 186–203, 2014.
[10] M. Drwal, E. Gerding, S. Stein, K. Hayakawa, and H. Kitaoka, "Adaptive pricing mechanisms for on-demand mobility," in *Proc. 16th Conf. Auto. Agents MultiAgent Syst.*, 2017, pp. 1017–1025.
[11] D. P. Bertsekas, "Dynamic programming and optimal control," *J. Oper. Res. Soc.*, vol. 47, no. 6, p. 833, 1996.
[12] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Hoboken, NJ, USA: Wiley, 2007.
[13] L. He, Z. Hu, and M. Zhang, "Robust repositioning for vehicle sharing," *Manuf. Service Oper. Manag.*, vol. 22, no. 2, pp. 241–256, Mar. 2020.
[14] T. Chen, B. Zhang, H. Pourbabak, A. Kavousi-Fard, and W. Su, "Optimal routing and charging of an electric vehicle fleet for high-efficiency dynamic transit systems," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3563–3572, Jul. 2018.
[15] M. Ammous, S. Belakaria, S. Sorour, and A. Abdel-Rahim, "Optimal cloud-based routing with in-route charging of mobility-on-demand electric vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2510–2522, Jul. 2019.
[16] S. Belakaria, M. Ammous, L. Smith, S. Sorour, and A. Abdel-Rahim, "Multi-class management with sub-class service for autonomous electric mobility on-demand systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7155–7159, Jul. 2019.
[17] C. J. R. Sheppard, G. S. Bauer, B. F. Gerke, J. B. Greenblatt, A. T. Jenn, and A. R. Gopal, "Joint optimization scheme for the planning and operations of shared autonomous electric vehicle fleets serving mobility on demand," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 6, pp. 579–597, Jun. 2019.
[18] E. S. Rigas, S. D. Ramchurn, and N. Bassiliades, "Algorithms for electric vehicle scheduling in large-scale mobility-on-demand schemes," *Artif. Intell.*, vol. 262, pp. 248–278, Sep. 2018.
[19] G. S. Bauer, J. B. Greenblatt, and B. F. Gerke, "Cost, energy, and environmental impact of automated electric taxi fleets in Manhattan," *Environ. Sci. Technol.*, vol. 52, no. 8, pp. 4920–4928, Apr. 2018.
[20] D. J. Fagnant and K. M. Kockelman, "The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios," *Transp. Res. C, Emerg. Technol.*, vol. 40, pp. 1–13, Mar. 2014.
[21] T. D. Chen, K. M. Kockelman, and J. P. Hanna, "Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions," *Transp. Res. A, Policy Pract.*, vol. 94, pp. 243–254, Dec. 2016.
[22] B. Jager, F. M. M. Agua, and M. Lienkamp, "Agent-based simulation of a shared, autonomous and electric on-demand mobility solution," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 250–255.

[23] N. Tucker, B. Turan, and M. Alizadeh, "Online charge scheduling for electric vehicles in autonomous mobility on demand fleets," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 226–231.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[25] F. Schneider, U. W. Thonemann, and D. Klabjan, "Optimization of battery charging and purchasing at electric vehicle battery swap stations," *Transp. Sci.*, vol. 52, no. 5, pp. 1211–1234, 2017.
[26] X.-R. Cao, "Stochastic learning and optimization—A sensitivity-based approach," *IFAC Proc. Volumes*, vol. 41, no. 2, pp. 3480–3492, 2008.
[27] *The Annual Traffic Census-2018*. Accessed: Jun. 13, 2020. [Online]. Available: https://www.td.gov.hk/filemanager/en/content_4953/annual%20traffic%20census%202018.pdf
[28] *Travel Characteristics Survey 2011 Final Report*. Accessed: Jun. 12, 2020. [Online]. Available: https://www.td.gov.hk/filemanager/en/content_4652/tcs2011_eng.pdf
[29] *LCQ18: Charging Facilities for Electric Vehicles*. Accessed: Jun. 12, 2020. [Online]. Available: https://www.epd.gov.hk/epd/sites/default/files/epd/english/environmentinhk/air/prob_solutions/files/EV_Charger_Locations_eng_01.xls
[30] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Nashua, NH, USA: Athena Scientific, 2019.

Liang Ni received the B.E. degree in information engineering from Zhejiang University, Hangzhou, China, in June 2014, the master's degree in electrical and computer engineering from Ohio State University, Columbus, OH, USA, in December 2015, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, in 2021. His research interests are on discrete resource allocation, stochastic modeling, and machine learning, with applications to transportation systems and energy systems.

Bo Sun (Member, IEEE) received the B.E. degree from the Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree from The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, in 2018. He is a Post-Doctoral Fellow at The Chinese University of Hong Kong. His research focuses on optimization and decision making under uncertainty with applications to real world networked systems.



Xiaoqi Tan received the Ph.D. degree from HKUST in 2018. During his Ph.D., he was a Visiting Research Fellow at the School of Engineering and Applied Science, Harvard University. He is an Assistant Professor at the Department of Computing Science, University of Alberta, and a fellow of the Alberta Machine Intelligence Institute (Amii), Edmonton, Canada. Prior to July 2021, he was a Post-Doctoral Fellow at the University of Toronto. His research interests span various topics in optimization and decision-making under uncertainty, with a focus on exploring the interplay between algorithms, learning, and incentives in online and multi-agent decision-making. On the practical side, his research is driven by optimization and decision-making problems in smart grid, electricity markets, transportation systems, and cloud computing.

Danny H. K. Tsang (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Moore School of Electrical Engineering, University of Pennsylvania, USA, in 1989. Upon graduation, he joined the Department of Computer Science, Dalhousie University, Canada. Later, he joined the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology (HKUST), in 1992, where he is currently a Professor of the department. He has been serving as a Leader for the Internet of Things Thrust Area, HKUST (Guangzhou Campus), since 2020. During his leave from HKUST, from 2000 to 2001, he assumed the role of Principal Architect at Sycamore Networks, USA. He was responsible for the network architecture design of Ethernet MAN/WAN over SONET/DWDM networks. He invented the 64B/65B encoding (U.S. Patent No.: US 6,952,405 B2) and contributed it to the proposal for Transparent GFP in the T1X1.5 standard that was advanced to become the ITU G.GFP standard. The coding scheme has been currently adopted by the International Telecommunication Union (ITU)'s Generic Framing Procedure recommendation GFP-T (ITU-T G.7041/Y.1303) and Interfaces for the Optical Transport Network (ITU-T G.709). His current research interests include cloud computing, edge computing, NOMA networks, and smart grids. He was nominated to become an IEEE fellow in 2012 and an HKIE fellow in 2013. He was a Guest Editor of IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS special issue on Advances in P2P Streaming Systems, an Associate Editor of *Journal of Optical Networking* published by the Optical Society of America, and a Guest Editor of IEEE SYSTEMS JOURNAL. He currently serves as a Technical Editor for *IEEE Communications Magazine*.