A Multi-Timescale and Bilevel Coordination Approach for Matching Uncertain Wind Supply With EV Charging Demand

Qilong Huang, *Student Member, IEEE*, Qing-Shan Jia, *Senior Member, IEEE*, and Xiaohong Guan, *Fellow, IEEE*

***Abstract*— The matching between random wind supply and electric vehicle (EV) charging demand can reduce the requirement of traditional power sources and the emission of CO_2. This problem is of great practical interest but involves system dynamics in multiple timescales. We consider this an important problem in this paper. In order to capture the randomness in the wind supply and EV charging demand, we formulate the problem as a bilevel Markov decision process. At the upper level, the charging demand of EVs in different locations is aggregated into multiple aggregators. The system operator dispatches power among the aggregators in a coarse timescale to maximize the wind power utilization. At the lower level, the aggregator schedules the charging process of individual EVs at a finer timescale to minimize the charging cost. In order to solve this large-scale problem, a bilevel simulation-based policy improvement (SBPI) method is developed. It is mathematically proved that the SBPI can improve from base policies in both levels. The performance of this multi-timescale and bilevel coordination approach is demonstrated through case studies in the city of Beijing.**

***Note to Practitioners*—This paper was motivated by the problem of reducing carbon emission and increasing utilization of renewable energy in power system. Being clean in terms of CO_2 emission, the wind power and EVs are becoming important parts in the supply and demand side. It is of great practical interest to schedule the (partially controllable) EV charging demand to match the random (uncontrollable) wind power generation. A key challenge to this problem is the coordination between the decision making at different timescales. In the coarse timescale (for example, the decision is made every hour), the forecasting of wind power generation contains a relatively large noise, but the prices of electricity from regular power plants are relatively cheap. In the finer timescale (for example, the decision is made every 15 min), because of the real-time information, the forecasting of wind power generation becomes more accurate, but the prices of electricity are relatively expensive. In this paper, we propose a multi-timescale scheduling model to address this issue. After formulating the problem as a Markov decision process, a bilevel simulation-based policy improvement method is developed to improve from base policies, which can be obtained by heuristics or existing policies in practice. Case studies in the city of Beijing show that the utilization of the wind power can be increased and the charging cost of the EVs can be reduced.**

***Index Terms*— Discrete event dynamic systems, electric vehicle (EV), Markov decision process (MDP), simulation-based policy improvement (SBPI), state aggregation, wind power.**

I. Introduction

THE environmental pollution and the shortage of fossil fuel have encouraged the development of renewable energy resources and the reduction of energy consumption. In such an effort, the wind power and the electric vehicle (EV) are good examples from the supply side and the demand side, respectively. For example, on the supply side in China, the wind energy installation has become the largest among all renewable energy by 2014. The installed capacity is about five Three Gorges Dam [1]. On the demand side, there will be 5 million EVs on the road in China by 2020 [2]. Therefore, it is of great practical interest to match the uncertain wind supply with EV charging demand.

This matching problem is usually related to two and probably more timescales. In the coarse of time scale (for example, the decision is made every hour), the wind power forecasting contains relatively large noise. But the electricity prices from the regular power plants are cheap. In the finer timescale (for example, the decision is made every 15 min), due to the availability of more information, the wind power forecasting becomes more accurate. As the energy procured in the finer timescale is usually provided by some expensive generators that have short start-up time, the electricity prices become more expensive [3]. Therefore, the coordination between the charging decisions at the two timescales should consider both the availability of the information and the price of electricity. We need to schedule the (partially controllable) EV charging demand to match the random (uncontrollable) wind power generation. This problem is challenging due to the following difficulties.

Manuscript received December 10, 2015; revised April 15, 2016; accepted June 1, 2016. Date of publication July 27, 2016; date of current version April 5, 2017. This paper was recommended for publication by Editor J. Wen upon evaluation of the reviewers' comments. This work was supported in part by the National Key Technology R&D Program under Grant 2013BAG18B00, in part by the 111 International Collaboration Program of China under Grant B06002, in part by the TNList Funding for Excellent Young Scholar, in part by the TNList Funding for Cross Disciplinary Research, in part by the Program for New Star of Science and Technology in Beijing under Grant xx2014B056, and in part by the National Natural Science Foundation of China under Grant 61174072, Grant U1301254, Grant 91224008, Grant 61222302, and Grant 61221063, and the Key R&D Project of China (Nos. 2016YFB0901901, 2016YFB0901903, and 2016YFB0901905). *(Corresponding author: Qing-Shan Jia.)*

Q. Huang and Q.-S. Jia are with the Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: hqlmailbox@gmail.com; jiaqs@tsinghua.edu.cn).

X. Guan is with the Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China, and also with the MOE KLINNS Laboratory, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xhguan@xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2016.2585180

First, multistage decision making. Regardless of how the EVs are charged during parking, the car should be ready to go when the user requires. Therefore, the charging of EV is coupled in time. The current decision making should also consider the cost in the future. Second, the uncertainties in the supply and demand. On the supply side, the accuracy of wind power forecasting degrades as the length of period increases. On the demand side, though the charging of an EV within a parking event may be scheduled (therefore controllable), the beginning and ending of a parking event are determined by the driver (and therefore uncontrollable). Third, the curse of dimensionality. The size of the state space and action space increases exponentially fast with respect to the number of EVs. There are usually hundreds of thousands of EVs in a megacity, such as Beijing. Therefore, a computationally feasible algorithm to schedule the charging of EVs is in demand.

We consider this important problem in this paper and make the following major contributions. First, this scheduling problem is formulated as a bilevel Markov decision process (MDP) to match the EV charging load with the wind power. At the upper level, the system operator dispatches power among the aggregators in a coarse timescale (for example, 1 h) to maximize the wind power utilization. At the lower level, the aggregator controls the charging process of individual EVs at a finer timescale (for example 15 min) to reduce the charging cost. Second, in order to effectively solve this large-scale charging problem, a bilevel simulation-based policy improvement (SBPI) method is developed. As each state in the upper level is an aggregated state from the lower level, this property is used to prove that the bilevel SBPI method can obtain improved policies from the given base policies for the lower level and upper level, respectively. Third, based on the real wind speed data and vehicle driving data in Beijing, it is numerically demonstrated that the proposed model and method can reduce the charging cost and improve the wind power utilization.

The rest of this paper is organized as follows. We review the literature in Section II, formulate the problem in Section III, present the solution methodology in Section IV, discuss the numerical results in Section V, and conclude in Section VI.

II. Literature Review

As the parking time of the EV is usually longer than its driving time, it can be considered as a large moving battery. Many works have been done to control the EV charging load to achieve different objectives, such as charging cost minimization [4], valley filling [5], and providing frequency regulation [6]. The existing works usually focus on the charging scheduling of all the individual EVs, which is a single-level control architecture. The control policy can be obtained based on the heuristic methods [7] or optimization methods [8], [9]. The heuristic method usually utilizes the state of charge (SOC) and the remaining parking time of EVs to prioritize EV charging decisions. The optimization method usually formulates this charging problem as a deterministic mixed integer programming to obtain the optimal charging policy. However, it is usually difficult to obtain the driving plans of EV users before optimization, as the optimization is usually implemented day-ahead [10]. Therefore, some works also formulate this charging problem as a stochastic programming model to handle the uncertainties in the EV moving [11].

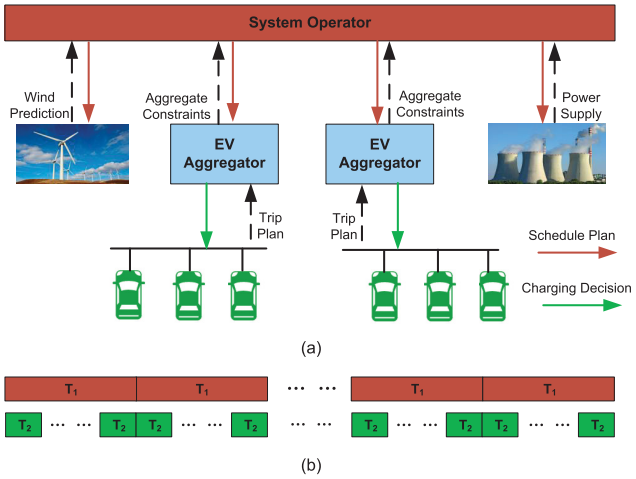

Fig. 1. (a) System architecture of the bilevel and multi-timescale scheduling problem. (b) Timescales for the upper level and lower level decisions.

As introduced above, the aforementioned works control the EV charging load on a single-level framework. When there are a large number of EVs, it is intractable to allow each EV user to directly join in the electricity market to procure the wind energy and conventional energy due to huge control variables and heavy communication burden [12]. Therefore, it is necessary to propose a bilevel framework for the EV load scheduling [13]–[15]. The upper level focuses on the scheduling of each EV aggregator (EVA), while the lower level focuses on the charging control of each EV in the aggregator. In most of these works, the timescales for the lower level and upper level are the same. However, the scheduling of conventional energy is usually repeated hourly due to the slow response of the thermal plants [16], while the charging control of each EV can be carried out in minute level. Therefore, we should schedule the EVAs in the upper level at a coarse timescale and control the EVs in each aggregator at a finer timescale [17].

There are mainly two advantages of this multi-timescale and bilevel framework. First, as the prediction of the wind power and EV charging load in the lower level (finer timescale) is more accurate than the prediction accuracy in the upper level (coarse timescale) [18], the lower level can adjust the charging process of EVs based on upper level decisions to reduce the deviation impact between the actual wind power in the lower level and predicted wind power in the upper level. Second, as the EVA collects all the charging requirements of EVs in the lower level, the control policy for the EVA in the upper level can be constructed based on the existing control policy in the lower level and *vice versa*. If we improve the control policy in one level, the improved control policy in the other level can be quickly constructed based on this property.

Based on the discussions above, we consider the system architecture, as shown in Fig. 1. The EVA collects the charging requirements of all its EVs and passes them to the system operator for the power dispatch. The existence of the EVA can save a lot of computation and communication burden [19].

At the upper level, the system operator focuses on a coarse timescale scheduling to match the supply and demand in the grid. It determines the optimal schedule plan for the wind power generators, the thermal plants, and the EVAs. At the lower level, each EVA focuses on a finer timescale scheduling to satisfy the charging demand of its individual EVs based on the upper level decisions. The uncertainties in this architecture mainly come from the wind power prediction and the driving plan prediction.

In the following, we will use MDP to formulate this problem. The advantages of MDP lie in two aspects [20]: 1) it is an effective tool to solve the stochastic programming problem with various approximate solution methods and 2) it can comprise nonlinear constraints and binary variables.

III. Problem Formulation

A. Assumptions

Before introducing the detailed MDP model for this problem, the following assumptions are made.

1) The charging power for each EV is set to be constant.
2) There is enough power supply from the thermal plant. Assumption 1 is the current practice as the constant charging power can prolong the service time of the battery [8]. Assumption 2 is used to ensure load balance when the wind energy cannot satisfy the charging demand of EVs.

B. Upper Level MDP Model

1) State: The system state is defined as $X_m = [W_m, N_m^k, E_m^k]$, where $m = m_0, m_0+1, \ldots, M$ denotes the stage index, $k = 1, 2, \ldots, K$ denotes the index for the EVA, W_m is the wind power at stage m, and N_m^k and E_m^k denote the number and total energy of parked EVs for the kth aggregator at time m. The current decision epoch is m_0 and the last decision epoch is M.

The wind power W_m can be calculated by using the following equation:

$$W_m = \begin{cases} W_{\text{cap}}, & v_{\text{rated}} < v_m \le v_{\text{cutout}} \\ W_{\text{cap}} \left(\dfrac{v_m}{v_{\text{rated}}} \right)^3, & v_{\text{cutin}} \le v_m \le v_{\text{rated}} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where v_m denotes the wind speed at stage m, v_{cutin} denotes the cut-in speed, v_{cutout} denotes the cut-out speed, v_{rated} denotes the rated speed, and W_{cap} denotes the wind capacity.

2) Action: The control action at the upper level is $A_m^U = [P_m^1, P_m^2, \ldots, P_m^K]$, where P_m^k denotes the dispatched power for the kth aggregator as defined in the following:

$$P_m^k = N_m^{k,c} \cdot P \tag{2}$$

$$P_m^{\text{EV}} = \sum_{k=1}^{K} P_m^k \tag{3}$$

where P denotes the constant charging power for the EV, $N_m^{k,c}$ denotes the number of EVs to be charged in the kth aggregator at time m, and P_m^{EV} denotes the total power of the EV charging load at stage m.

3) System Dynamics: Given the system state X_m and action A_m^U, the system dynamics are depicted as follows:

$$N_{m+1}^k = N_m^k + N_{m+1}^{k,\text{in}} - N_{m+1}^{k,\text{out}} \tag{4}$$

$$E_{m+1}^k = E_m^k + P_m^k \cdot T_U + E_{m+1}^{k,\text{in}} - E_{m+1}^{k,\text{out}} \tag{5}$$

where T_U is the time interval between adjacent decision epoch, $N_{m+1}^{k,\text{in}}$ and $N_{m+1}^{k,\text{out}}$ are the number of newly arrival and departure EVs at the beginning of time $m+1$, and $E_{m+1}^{k,\text{in}}$ and $E_{m+1}^{k,\text{out}}$ are the total energy of newly arrival and departure EVs at the beginning of time $m+1$.

4) Constraints: For each EVA, the following constraints should be satisfied:

$$0 \le N_m^{k,c} \le N_m^k \tag{6}$$

$$\delta_{\min}^k \cdot N_m^k \cdot E_{\text{cap}} \le E_m^k + P_m^k \cdot T_U \le \delta_{\max}^k \cdot N_m^k \cdot E_{\text{cap}} \tag{7}$$

$$0 \le N_{m+1}^{k,\text{in}} \le N_{\text{reg}}^k - N_m^k$$

$$0 \le E_{m+1}^{k,\text{in}} \le N_{m+1}^{k,\text{in}} \cdot E_{\text{cap}} \tag{8}$$

$$0 \le N_{m+1}^{k,\text{out}} \le N_m^k$$

$$0 \le E_{m+1}^{k,\text{out}} \le N_{m+1}^{k,\text{out}} \cdot E_{\text{cap}} \tag{9}$$

where E_{cap} denotes the battery capacity of the EV, $\delta_{\min}^k$ and $\delta_{\max}^k$ denote the minimum and maximum average SOC of the kth aggregator, and N_{reg}^k denotes the number of registered EVs in the kth aggregator. Constraint (6) represents that the number of EVs to be charged should be smaller than the number of current parked EVs. Constraint (7) regulates the bound of the energy level to ensure the driving needs of EVs from a point of statistical view. Constraints (8) and (9) regulate the bound for the number and total energy of the arrival and departure EVs, respectively. These two constraints are used for the simulation of the EV arrival and departure events in the future.

5) Objective Function: As the responsibility of the system operator is to improve the wind power penetration while ensure load balance, the optimization of the matching degree is considered in the objective function. Meanwhile, in order to incent the participation of the EV drivers, the charging cost should also be considered. Therefore, the one-step cost function is defined as follows:

$$C_m^U(X_m, A_m^U) = (1-\lambda) \cdot \frac{\left| W_m - P_m^{\text{EV}} \right|}{\max\left(W_m, P_m^{\text{EV}} \right)} + \lambda \cdot \beta_m \cdot \max\left(P_m^{\text{EV}} - W_m, 0 \right) \cdot T_U \tag{10}$$

where $P_m^G = \max(P_m^{\text{EV}} - W_m, 0)$ is the required thermal power to balance the demand, β_m is the electricity price, and λ is the weighting parameter. The first part in (10) denotes the normalized mismatch gap between the EV charging load and the wind power. The small mismatch gap means the large wind power utilization for EV charging. The second part denotes the total charging cost. Based on our experience [21], λ can be set as $3/\max(W_{\text{cap}}, N_{\text{reg}} P)$ to rescale these two parts into commeasurable values, where $N_{\text{reg}} = \sum_{k=1}^{K} N_{\text{reg}}^k$.

The expected total cost is used as the objective function in the upper level MDP model, which is shown in the following:

$$J^U(\mu, X_{m_0}) = \mathbf{E}_\mu \left[\sum_{m=m_0}^{M} \left\{ (1-\lambda) \cdot \frac{\left| W_m - P_m^{\mathrm{EV}} \right|}{\max\left(W_m, P_m^{\mathrm{EV}}\right)} + \lambda \cdot \beta_m \cdot \max\left(P_m^{\mathrm{EV}} - W_m, 0\right) \cdot T_U \right\} \right] \quad (11)$$

where X_{m_0} is the initial state and μ is the scheduling policy. The policy μ decides the specific action for the state, that is

$$\mu(X_m) = A_m^U \quad \forall m = m_0, \quad m_0 + 1, \ldots, M. \quad (12)$$

The objective is to find an optimal charging policy μ^* that minimizes the expected total cost (11), that is

$$\min_{\mu \in \Omega} J^U(\mu, X_{m_0}) \quad (13)$$

where Ω is the policy space of the upper level MDP model.

According to Fig. 1, at stage m_0, the system operator will solve this upper level MDP model and dispatch power $P_{m_0}^k$ to each EVA based on the policy μ^*.

C. Lower Level MDP Model

The lower level optimization is carried out in the aggregator level. Each EVA independently controls the charging process of its EVs. Due to the finer timescale in the lower level decision, the wind power prediction and the EV moving prediction are more accurate. Therefore, the lower level should determine the optimal charging process for each EV based on the dispatch power $P_{m_0}^k$ and improved prediction information. The details of this lower level MDP model are shown in the following.

1) State: The system state for the kth aggregator in the lower level is defined as $S_n(k) = [W_n, L_n^{i(k)}, E_n^{i(k)}]$, where $n = n_0, n_0+1, \ldots, N$ is the stage index, $i(k) = 1, 2, \ldots, N_{\mathrm{reg}}^k$ is the EV index for the kth aggregator, $L_n^{i(k)}$ and $E_n^{i(k)}$ are the remaining parking time and remaining required charging energy for the ith EV in the kth aggregator at stage n, and W_n is the wind power at stage n. $L_t^{i(k)}$ and $E_t^{i(k)}$ are defined as follows:

$$\begin{cases} \left(L_n^{i(k)} > 0, E_n^{i(k)} \geq 0\right), & \text{if } I_{i(k)}(n) = 1 \\ \left(L_n^{i(k)} = 0, E_n^{i(k)} = 0\right), & \text{if } I_{i(k)}(n) = 0 \end{cases} \quad (14)$$

where $I_{i(k)}(n) = 1$ if the ith EV is parked at stage n, otherwise $I_{i(k)}(n) = 0$. The current decision epoch is n_0 and the last decision epoch is N.

2) Action: The control action for the kth aggregator at stage n is $A_n^L(k) = [z_n^1, z_n^2, \ldots, z_n^{N_{\mathrm{reg}}^k}]$, where $z_n^{i(k)}$, $i = 1, 2, \ldots, N_{\mathrm{reg}}^k$ is a binary variable that satisfies

$$z_n^{i(k)} = \begin{cases} 1 \text{ or } 0, & \text{if } I_{i(k)}(n) = 1 \\ 0, & \text{if } I_{i(k)}(n) = 0. \end{cases} \quad (15)$$

$z_n^{i(k)} = 1$ if the ith EV is selected to be charged at stage n, otherwise $z_n^{i(k)} = 0$. Thus, based on Assumption 1, the charging energy for the ith EV at stage n is $z_n^{i(k)} \cdot P \cdot T_L$, where T_L is the time interval between adjacent decision epoch in the lower level.

3) System Dynamics: Given the system state $S_n(k)$ and action $A_n^L(k)$, the system dynamics are depicted as follows:

$$L_{n+1}^{i(k)} = \begin{cases} L_n^{i(k)} - T_L, & \text{if } I_{i(k)}(n) = 1 \\ \tau_{n+1}^{i(k)}, & \text{if } I_{i(k)}(n) = 0, \quad I_{i(k)}(n+1) = 1 \\ 0, & \text{if } I_{i(k)}(n) = 0, \quad I_{i(k)}(n+1) = 0 \end{cases} \quad (16)$$

$$E_{n+1}^{i(k)} = \begin{cases} E_n^{i(k)} - z_n^{i(k)} P T_L, & \text{if } I_{i(k)}(n) = 1 \\ \eta_{n+1}^{i(k)}, & \text{if } I_{i(k)}(n) = 0, \quad I_{i(k)}(n+1) = 1 \\ 0, & \text{if } I_{i(k)}(n) = 0, \quad I_{i(k)}(n+1) = 0 \end{cases} \quad (17)$$

where $\tau_{n+1}^{i(k)}$ and $\eta_{n+1}^{i(k)}$ are both random variables, which denote the parking time and the required charging energy for the ith EV at stage $n+1$, if the ith EV is driving at stage n and begins to park at stage $n+1$, respectively.

4) Constraints: The action $A_n^L(k)$ at stage n corresponding to the state $S_n(k)$ is constrained by the following constraints:

$$P_n^k = P \cdot \sum_{i=1}^{N_{\mathrm{reg}}^k} z_n^{i(k)} \quad (18)$$

$$0 \leq E_n^{i(k)} \leq E_{\mathrm{cap}} \quad (19)$$

$$0 \leq E_n^{i(k)} \leq P \cdot L_n^{i(k)}. \quad (20)$$

Constraint (18) represents the total charging power of the kth aggregator. Constraint (19) denotes that the required charging energy should not exceed the battery capacity. Constraint (20) regulates that the remaining required charging energy at stage n should not exceed the maximum charging energy that can be provided during the remaining parking time. When the EV departs, $L_n^{i(k)}$ will become zero which will require that $E_n^{i(k)}$ should also be zero. This ensures that the charging demand is satisfied when the EV departs.

5) Objective Function: As the EV drivers in the aggregator care about the charging cost, the responsibility of each aggregator is to minimize the charging cost of its EVs. Therefore, the one-step cost function is defined as follows:

$$\text{if } \frac{T_U}{T_L}(m_0 - 1) + 1 \leq n \leq \frac{T_U}{T_L} m_0$$

$$C_n^L\left(S_n(k), A_n^L(k)\right) = \beta_n T_L \cdot \max\left(P_n^k - \left(W_n + P_{m_0}^G\right) \frac{P_{m_0}^k}{P_{m_0}^{\mathrm{EV}}}, 0 \right)$$

$$\text{if } n > \frac{T_U}{T_L} m_0, \quad C_n^L\left(S_n(k), A_n^L(k)\right) = \beta_n T_L P_n^k \quad (21)$$

where β_n is the electricity price at stage n. Note that as $(n_0 - 1)T_L \geq (m_0 - 1)T_U$, there is no possibility that $n < (T_U/T_L)(m_0 - 1) + 1$. The first condition in (21) indicates that the decision at stage n in the lower level is still influenced by the upper level decision at stage m_0. Therefore, the charging cost is generated when the thermal power and wind power provided by the system operator cannot satisfy the actual demand of EVs in the lower level. The second condition in (21) indicates that the upper level has not made any decision yet at stage n. Therefore, the charging cost is totally afforded

by the lower level. $P_{m_0}^k / P_{m_0}^{\mathrm{EV}}$ is the power dispatch ratio that means the power dispatch among aggregators is proportional to the estimated total charging power of each aggregator in the upper level.

The expected total cost is also used as the objective function in the lower level, which is shown in the following:

$$J^L(\pi, S_{n_0}(k)) = \mathbf{E}_\pi \left[\sum_{n=n_0}^{N} C_n^L\big(S_n(k), A_n^L(k)\big) \right] \tag{22}$$

where S_{n_0} is the initial state and π is the scheduling policy that satisfies $\pi(S_n(k)) = A_n^L(k)$.

The objective for the kth aggregator is to find an optimal charging policy π^* that minimizes the expected total cost (22), that is

$$\min_{\pi \in \Pi} J^L(\pi, S_{n_0}(k)) \tag{23}$$

where Π is the policy space of the lower level MDP model.

Due to the uncertainties and large number of EVs in this bilevel MDP, it will face to the curse of dimensionality to find the optimal policies for the upper level and lower level. Therefore, in Section IV, we will introduce an approximate solution methodology to solve this problem.

IV. Solution Methodology

In this section, we will introduce a bilevel SBPI method to approximately solve this problem. We will first introduce the traditional SBPI, then introduce the details of the bilevel SBPI, and prove the cost improvement property of the bilevel SBPI in the end.

A. Traditional Simulation-Based Policy Improvement

The main idea of the SBPI (also called rollout) is to obtain an improved policy from the given rule-based or heuristic-based policy [17], [22], [23]. The SBPI updates the action only for the current state, which can avoid traversing the whole state space. The SBPI is an approximate solution method [24]. Compared with approximate dynamic programming method, the SBPI uses the base policy to estimate the optimal value function, which is less time-consuming and challenging [25]. In the SBPI, the Q-factor is used to evaluate the performance of each state-action (S_t, A_t) pair, that is

$$Q_t(S_t, A_t) = C_t(S_t, A_t) + \mathbf{E}[V_{t+1}(S_{t+1}) | S_t, A_t] \tag{24}$$

where C_t is the one-step cost function at stage t and $V_{t+1}(S_{t+1})$ is the optimal value function of state S_{t+1} at stage $t+1$, which can be depicted as follows:

$$V_{t+1}(S_{t+1}) = \mathbf{E}\left[\sum_{j=t+1}^{T} C_j(S_j, \pi^*(S_j)) | S_{t+1}) \right] \tag{25}$$

where T is the last stage and π^* is the optimal policy. The optimal action regarding S_t is

$$A_t^* = \arg\min_{A_t \in \mathcal{A}} Q_t(S_t, A_t) \tag{26}$$

where $\mathcal{A}$ is the action space.

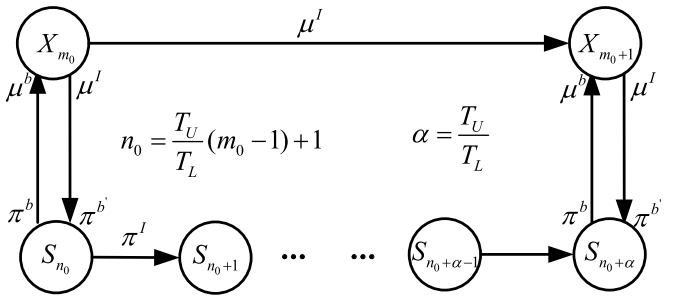

Fig. 2. Flowchart of the bilevel SBPI.

However, the optimal policy in $V_{t+1}(S_{t+1})$ is not known *a priori* (otherwise, we have already solved this problem). As there exist many (rule-based or heuristic-based) base policies in practice, the SBPI uses the base policy and simulations to approximate the optimal value function and its expectation. Therefore, the Q-factor is approximated by

$$\hat{Q}_t(S_t, A_t) = C_t(S_t, A_t) + \frac{1}{M_a} \sum_{m=1}^{M_a} \sum_{j=t+1}^{T} [C_j(S_j, \pi^b(S_j)) | \zeta_m] \tag{27}$$

where M_a denotes the number of sample paths, ζ_m denotes the randomness in the mth sample path, and π^b denotes the base policy. The SBPI chooses the action, which minimizes the approximate Q-factor, that is

$$\pi^I(S_t) = \arg\min_{A_t \in \mathcal{A}} \hat{Q}_t(S_t, A_t) \tag{28}$$

and π^I denotes the improved policy.

Note that the Q-factor of each action in (27) can be parallelly computed to reduce the computation time. The SBPI can be applied online and it can ensure $J(\pi^I, S_1) \leq J(\pi^b, S_1)$ [22].

B. Bilevel Simulation-Based Policy Improvement

As the EV load scheduling problem is formulated as a bilevel MDP, the traditional SBPI should be modified to take advantage of this bilevel architecture. In the following, we will call this modified SBPI as the bilevel SBPI.

In the bilevel architecture, the lower level collects the charging requests and passes them to the upper level. Then, the upper level makes the decision and passes it to the lower level. The bilevel SBPI is applied in a similar way. Fig. 2 shows this algorithm. Let $S_n = [S_n(1), \ldots, S_n(K)]$. Then, the upper level state X_m can be considered as the state aggregation of the lower level state S_n, i.e., $S_n \in X_m$. The algorithm is summarized as follows.

Step 1: At stage m_0, the EVA and the wind farm generate M_a future sample paths of EV parking events and wind power, respectively. The number and the energy of the arrival and departure EVs in the future will be computed based on these sample paths and passed to the system operator.

Step 2: The lower level chooses a base policy π^b. The upper level constructs the base policy μ^b based on π^b

in the lower level. The construction rule can be as follows, $\forall n = (m-1)T_U/T_L + 1$:

$$\mu^b(X_m) = P \cdot \left[\pi^b(S_n(1)) \cdot \mathbf{1}, \ldots, \pi_n^b(S_n(K)) \cdot \mathbf{1}\right] \tag{29}$$

where $\mathbf{1}$ is the column vector with unit elements.

Step 3: The upper level obtains an improved policy μ^I from the base policy μ^b by using SBPI.

Step 4: Then, the lower level updates the base policy (π^b to $\pi^{b'}$) by projecting μ^I to the lower level, i.e., $\forall k = 1, 2, \ldots, K, \forall n = n_0, n_0+1, \ldots, n_0+\alpha-1$

$$\pi^{b'}(S_n(k)) \cdot \mathbf{1} = \mu^I(X_{m_0})(k)/P \tag{30}$$

where $\mu^I(X_{m_0})(k)$ denotes the kth element in $\mu^I(X_{m_0})$. Equation (30) means that the kth aggregator in the lower level should select $\mu^I(X_{m_0})(k)/P$ EVs to charge. The selection rule for charging can be heuristic, e.g., randomly pick $\mu^I(X_{m_0})(k)/P$ EVs to charge or pick $\mu^I(X_{m_0})(k)/P$ EVs that have the lowest SOC or shortest remaining parking time to charge.

Step 5: Based on the updated base policy $\pi^{b'}$, the lower level obtains the improved policy π^I at each decision epoch between n_0 and $n_0+\alpha-1$ by using SBPI and the newly generated sample paths of wind power and EV parking events.

Step 6: Set $m_0 = m_0 + 1$, jump to Step 1.

Note that although several EVA need optimization in Step 5, the optimization for each EVA can be parallelly computed to reduce the computation time. Furthermore, the action space in the lower level can be largely reduced by implementing the following steps. First, group the parked EVs by their remaining parking time. Second, obey the following rule to select the EVs to charge in each group, i.e., pick the EVs by their remaining required charging energy in descent order. These two steps can reduce large amounts of suboptimal actions. The details of the SBPI for the lower level optimization can be found in this paper [23].

C. Cost Improvement Property of Bilevel SBPI

In this section, we will prove the cost improvement property of the bilevel SBPI, i.e., $J^L(\pi^I, S_{n_0}) \le J^L(\pi^b, S_{n_0})$. As introduced that the upper level state X_m is the state aggregation of the lower level state S_n, we will use this aggregation characteristic to demonstrate the cost improvement property. In the proof, we will assume that the one-step cost function for the lower level is $C(S_n, \pi(S_n)) = \beta_n T_L \cdot \max(\sum_{k=1}^K P_n^k - W_n, 0)$ and the cost function for the upper level is similar with P_n^k replaced by P_m^k and T_L replaced by T_U. $\forall m$ and n satisfy $n = (T_U/T_L)(m-1)+1$, let

$$\epsilon_m^u = \max_{S_n^i, S_n^j \in X_m^u} \left| J^L(\pi, S_n^i) - J^L(\pi, S_n^j) \right| \tag{31}$$

represents the maximal aggregation error in the aggregated state X_m^u, where the superscripts in X_m and S_n denote the specific states in the upper level and lower level state space, respectively. Let $\epsilon_m = \max_{\{X_m^u \in X_m\}} \epsilon_m^u$, then there are the following theorems.

Theorem 1: $\forall m, n$ satisfy $n = (T_U/T_L)(m-1)+1$ and the relationship between upper level policy μ and lower level policy π satisfies (29) or (30), then there is

$$\left| J^U(\mu, X_m^u) - J^L(\pi, S_n^l) \right| \le \sum_{\tau=m}^{M} \epsilon_\tau \quad \forall S_n^l \in X_m^u. \tag{32}$$

The proof for Theorem 1 is given in Appendix A. Theorem 1 shows the performance difference between the states in the lower level and their aggregated state in the upper level. Note that when $M \to \infty$ and the discounted factor γ is used, $\epsilon_m \to \epsilon$ and the bound can be reduced to $\epsilon/(1-\gamma)$, which is consistent with the bound in [26].

Theorem 2: Let $\delta_m = J^U(\mu^b, X_m) - J^U(\mu^I, X_m)$, $\epsilon_\tau^{\mathrm{LU}}$, and $\epsilon_\tau^{\mathrm{UL}}$ denote the aggregation error from lower level to upper level and from upper level to lower level, respectively, then there is

$$\begin{aligned} J^L(\pi^{b'}, S_n^l) - J^L(\pi^b, S_n^l) &\le \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{LU}} + \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{UL}} - \delta_m \\ J^L(\pi^{b'}, S_n^l) - J^L(\pi^b, S_n^l) &\ge -\left(\delta_m + \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{LU}} + \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{UL}}\right). \end{aligned} \tag{33}$$

Especially, when $\delta_m \ge \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{LU}} + \sum_{\tau=m}^{M} \epsilon_\tau^{\mathrm{UL}}$, there is $J^L(\pi^{b'}, S_n^l) \le J^L(\pi^b, S_n^l)$.

The proof for Theorem 2 is given in Appendix B. Theorem 2 shows the performance difference between the lower level policies π^b and $\pi^{b'}$. It is notable that it has been proved in [22] that $J^L(\pi^I, S_n^l) \le J^L(\pi^{b'}, S_n^l)$ and $J^U(\mu^I, X_m) \le J^U(\mu^b, X_m)$. Therefore, when the last condition in Theorem 2 is satisfied, it can ensure $J(\pi^I, S_{n_0}) \le J(\pi^b, S_{n_0})$.

In our problem, the aggregation error comes from the EVA as it aggregates the EV information and neglects the detailed charging requirement of each EV in the lower level. This causes the inconsistency between the lower level model and the upper level model. Fortunately, as the performance improvement δ_m of the upper level is usually significant, the cost improvement property of this bilevel SBPI still holds, which will be demonstrated in the numerical experiments.

V. Numerical Results

A. Parameter Settings

In this section, we will use the real wind data and vehicle data to demonstrate the effectiveness of our model and method. The wind speed at 12 m-height is collected at a weather station in Tsinghua University [27]. The Vestas V100-2 MW wind turbine is used [28] and its specification is shown in Table I. As the height of the wind turbine is 100 m, the wind speed at this height can be obtained by using the following equation:

$$V_{h_1} = V_{h_2}\left(\frac{h_1}{h_2}\right)^{\sigma} \tag{34}$$

where in our experiment $h_1 = 100$ m, $h_2 = 12$ m, and $\sigma = 1/7$ [29]. Based on the experiment settings in [16],

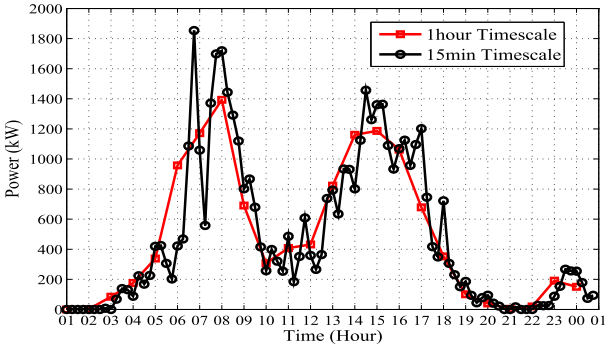

Fig. 3. Wind power at 1-h-timescale and 15-min-timescale.

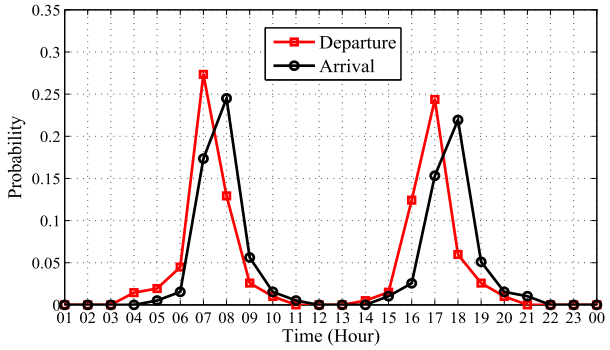

Fig. 4. Probability of the vehicle departure and arrival event.

TABLE I
PARAMETER SETTINGS OF THE WIND GENERATOR AND EV

Parameter	Setting	Parameter	Setting
E_{cap}	57kWh	W_{cap}	2MW
P	3kW	height	100m
ω	0.3kWh/km	v_{rated}	12m/s
T_U	1Hour	v_{cutout}	22m/s
T_L	15min	v_{cutin}	3m/s

we assume the wind power follows normal distribution. The mean value is the actual generated wind power at 1-h-timescale and 15-min-timescale, which is shown in Fig. 3. The standard deviation is 10% of the mean value.

The vehicle data in Beijing are used to simulate the parking event in the future [30]. Fig. 4 shows the probability of the vehicle arrival and departure event for going to work and going home. The trip distance is shown in Fig. 5 and the fitting function is the power-law function with the cutoff, which is shown in the following:

$$f(x_d) \propto x_d^{-a} e^{-bx_d} \tag{35}$$

where x_d is the trip distance and f is its probability distribution. The values of a and b are estimated as 0.5065 and 0.0735, respectively. Based on these probability distributions, we can generate the EV arrival time, departure time, and trip distance by randomly sampling. The required charging energy for each parking event can be computed by using the trip distance x_d and the electric drive efficiency w. The specification of the EV comes from BYD e6 [31] and is shown in Table I.

In the experiment, we will consider this scheduling problem on a daily basis. The timescales for the two levels are shown

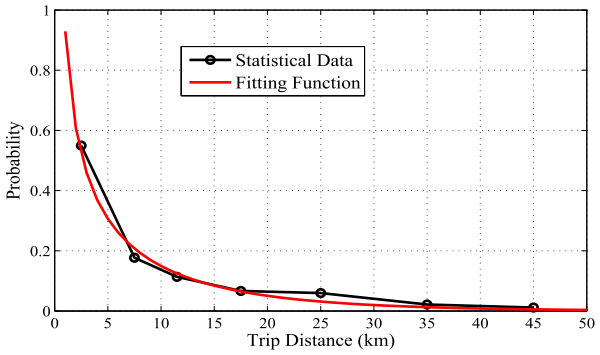

Fig. 5. Statistical data and fitting distribution of the trip distance.

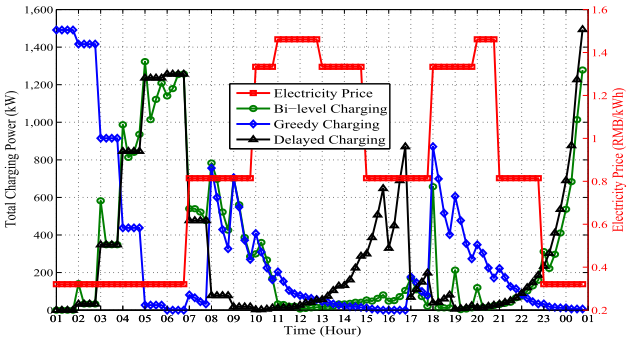

Fig. 6. Electricity price and total charging power in different charging policies.

in Table I and the electricity price for commercial usage in Beijing is shown in Fig. 6. $\delta^k_{\max}$ and $\delta^k_{\min}$ are set to be 1 and 0.25 for the upper level, respectively. We first conduct the experiment of five aggregators with 100 EVs in each, and then conduct the experiment of three aggregators with 1000 EVs in each. The action space is limited to 10 000 and $M_a = 50$ sample paths are used to estimate the Q-factor based on the above probability distributions. This experiment is run on a desktop with an Intel Core2 2.8-GHz CPU and 4-GB memory. The following three control policies are considered to compare.

1) Greedy Charging: Each EV will be charged to the full state as soon as possible. This is the current control strategy for the EV charging.
2) Delayed Charging: Each EV will be postponed to charge as long as possible. This policy can utilize the low electricity price at early morning and late night to save the charging cost.
3) Bilevel Charging: This is our proposed method. The charging policy is obtained by using the bilevel SBPI, where the greedy charging policy is chosen as the base policy.

B. Charging Cost and Matching Degree Analysis

The total charging power of these three control policies are shown in Fig. 7. In the experiment settings, each EV has a charging requirement at the initial stage and requires to finish all the charging requirements at the last stage. This causes the total charging power at the first stage and last stage are not the same. As can be seen in Fig. 7, the greedy

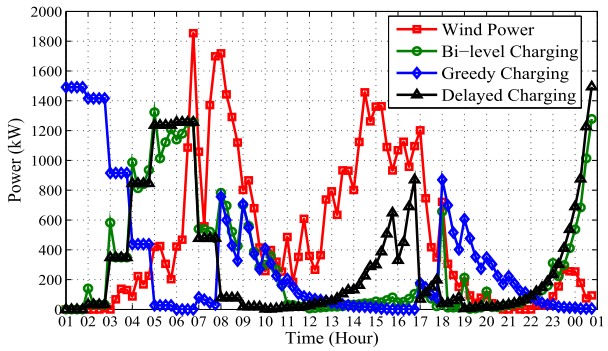

Fig. 7. Wind power and total charging power in different charging policies.

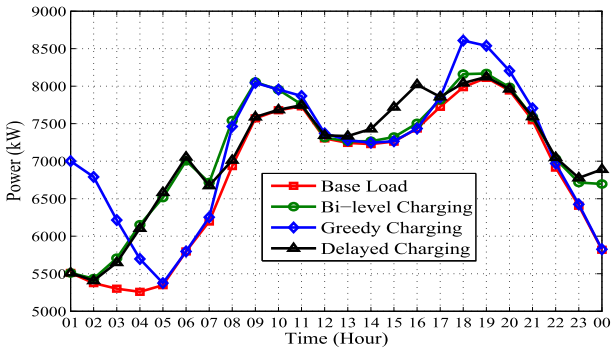

Fig. 8. Load patterns of the three charging policies.

policy performs worst during the early morning. The total charging power of the greedy policy keeps decreasing within this period, while the wind power keeps increasing during this period. In contrast, the delayed charging policy and the bilevel charging policy can make the EV charging power follow the wind power. Due to this reason, the total charging costs of the greedy policy, delayed charging policy, and bilevel charging policy are 2609.3, 1237.4, and 1623.2 RMB, respectively. Note that as the system operator not only focuses on the charging cost optimization, but also focuses on the matching degree optimization. This makes that the charging cost of the bilevel charging policy is not the smallest among the three control policies.

The main difference between the bilevel charging policy and delayed charging policy is within the time period (7:00–20:00). It can be found that the biggest variation of the wind power appears within the time period (7:00–9:00). The bilevel charging policy will make the charging power follow the wind power within (7:00–9:00) as much as possible, while the delayed charging policy does not consider the matching between the wind power and EV charging load. By defining the matching degree as $1 - |W_m - P_m^{\mathrm{EV}}| / \max(W_m, P_m^{\mathrm{EV}})$, the average matching degree of the greedy policy, delayed charging policy, and bilevel charging policy are 0.2083, 0.1758, and 0.3408, respectively. Due to this increased matching degree, the total spilled wind energy is reduced by 527.8 kWh in the bilevel charging policy compared with the delayed charging policy. This demonstrates that the bilevel charging policy can have a good balance between charging cost optimization and matching degree optimization.

Note that, the bilevel charging policy is obtained based on the greedy policy. It can be found that the improvement of the bilevel charging policy is significantly compared with the greedy charging policy. This demonstrates the effectiveness of the bilevel SBPI method.

C. Pricing and Load Shifting Effect

Fig. 6 shows the relationship between the electricity price and the total charging power in these control policies. It can be found that both the delayed charging policy and the bilevel charging policy have high charging power during time periods (1:00–6:00) and (23:00–1:00) when the electricity price is cheap, while the greedy policy has high charging power during time period (18:00–22:00) when the electricity price is expensive. Due to this reason, the average unit prices for charging the greedy charging policy, delayed charging policy, and bilevel charging policy are 0.3448, 0.1641, and 0.2146 RMB/kWh, respectively.

It is also important to analyze the impact of the EV charging load to the base load in the grid. Fig. 8 shows the total load of the three control policies compared with the base load. The base load is the rescaled load of a typical working day in Beijing (the rescaled factor is 1/1000). This rescaled operation can make the base load and the EV charging load comparable. From Fig. 8, it can be seen that all three policies can fill the valley of the base load. However, the greedy policy also largely increases the peak of the base load, while the other policies slightly increase the peak of the base load. In the base load, the peak-valley gap is 2852 kW. The greedy policy increases this gap to 3233.7 kW, while the delayed charging policy and the bilevel charging policy decrease this gap to 2716.5 and 2735.2 kW, respectively. This shows that the scheduling of EV charging load is helpful to the load shifting.

D. Inconsistency Analysis of the Bilevel Model

As each EVA should aggregate the charging demand of its EVs and pass it to the upper level, some information will be lost during this aggregation, such as the state of the individual EV. Therefore, it is necessary to analyze the inconsistency between the lower level model and the upper level model. Figs. 9 and 10 show the state transition of the number and total energy of parked EVs in each aggregator. These information can be accurately provided by each EVA. It can be found that most of the EVs are driven at 7:00 and 17:00 for going to work and going home. The total energy of parked EVs reaches its peak at 11:00 and 23:00. This shows that the charging demand of EVs can be satisfied before departure when the bilevel charging policy is implemented.

Fig. 11 shows the estimated charging power in the upper level and the actual charging power in the lower level. It can be found that the main differences exist in time periods (1:00–6:00) and (22:00–24:00). These differences are caused by the inconsistency between the upper level model and lower level model. During these periods, the number of EVs that needs to be charged is very large but the upper level has no information about the detailed charging demand of all the individual EVs. This causes the large gap of the charging

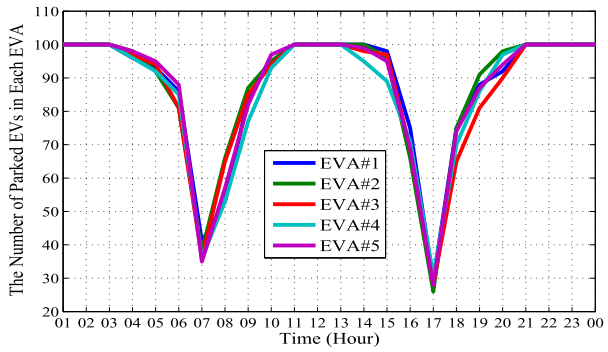

Fig. 9. Number of parked EVs in each EVA.

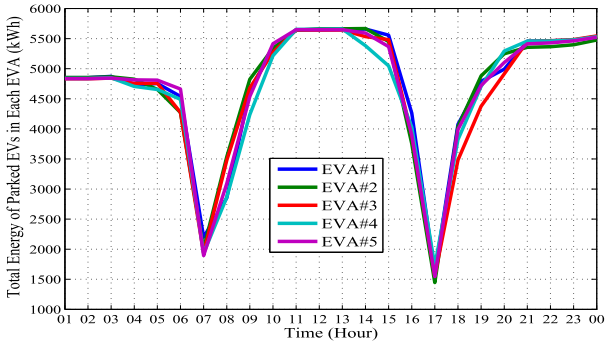

Fig. 10. Total energy of parked EVs in each EVA.

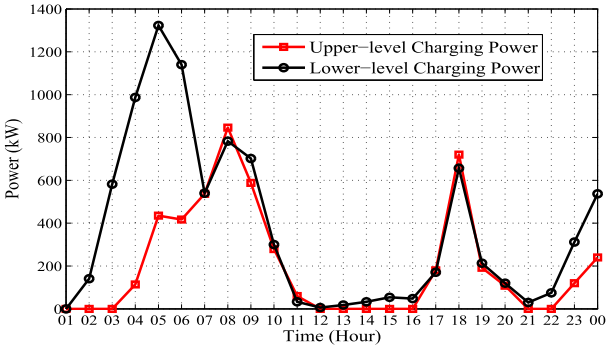

Fig. 11. Total charging power in the upper level and lower level with the minimum average SOC $\delta^k_{\min}$ set as 0.25.

power between the lower level and upper level. Due to this reason, the optimized matching degree in the upper level is reduced from 0.6319 to 0.3408 in the lower level. Therefore, it is important to change the value of $\delta^k_{\min}$ for each aggregator at each decision epoch based on the actual charging demand in the lower level. Fig. 12 shows the estimated charging power in the upper level when the minimum average SOC $\delta^k_{\min}$ is increased. In this case, $\delta^k_{\min}$ is set to be 0.85 during time period (2:00–6:00). Due to this increase of the minimum energy level, the gap between the upper level estimation and the lower level realization can be reduced. When using our method in practice, the value of $\delta^k_{\min}$ can be set based on the power gap between lower level and upper level in the previous day as the daily driving habits of EVs are similar.

E. Effectiveness of Multi-Timescale Charging Control

As the upper level and lower level have different timescales, we should compare the multi-timescale charging control policy

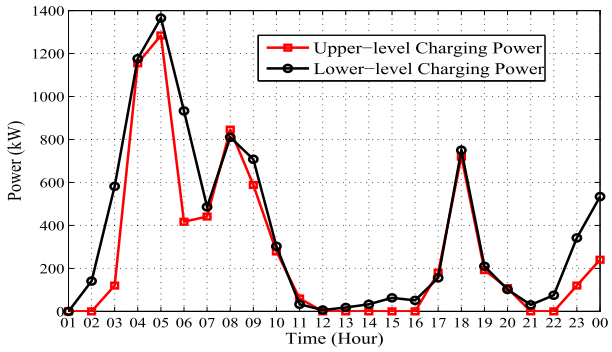

Fig. 12. Total charging power in the upper level and lower level with the increased minimum average SOC.

TABLE II
PERFORMANCE OF THE MULTI-TIMESCALE AND SINGLE-TIMESCALE CHARGING CONTROL POLICIES

Policies	Total Cost	MD	Peak-Valley Gap	AUP
Multi-timescale	1623.2RMB	0.3408	2735.2kW	0.2146RMB/kWh
Single-timescale	1741.2RMB	0.3499	3076.4kW	0.2309RMB/kWh

with the single-timescale control policy. The aforementioned greedy policy and delayed policy can be considered as the single-timescale control policies, which are applied in the lower level. Therefore, we additionally consider the single-timescale control policy applied in the upper level. This single-timescale policy makes the same upper level decisions as our method and follows Step 4 in our algorithm to heuristically dispatch power to all the EVs in the lower level. The performance of these single-timescale and multi-timescale control policies is shown in Table II, where MD represents average matching degree and AUP represents average unit price for charging. It can be seen that as there is no multistage optimization for the lower level in the single-timescale control policy, the total charging cost and the average charging price are increased compared with the multi-timescale control policy. This increased cost can be considered as the penalty for the short-sighted lower level decision making in the single-timescale control policy. Table II also shows that the peak-valley gap of the multi-timescale control policy is also smaller than the single-timescale control policy, while the matching degree of the former is slightly reduced compared with the latter. Therefore, it demonstrates that it is necessary to implement multi-timescale optimization for the upper level and lower level.

F. Scalability Analysis

One of the main advantages of the bilevel scheduling is that it can reduce the computation burden. Therefore, we conduct another experiment of three aggregators with 1000 EVs in each. With the bilevel architecture, the average running time for each decision epoch is about 12.9 min. In contrast, if no aggregator is used and the centralized optimization is used, the average running time for each decision epoch will be about 26.7 min [23]. This demonstrates that the bilevel scheduling can speed up the optimization. The performance of these three policies is shown in Table III. Note that as the

TABLE III
PERFORMANCE OF THE POLICIES FOR A LARGE-SCALE PROBLEM

Policies	Total Cost	MD	Peak-Valley Gap	AUP
Greedy	25117RMB	0.1992	8860.9kW	0.5496RMB/kWh
Delayed	15694RMB	0.3099	7931.5kW	0.3455RMB/kWh
Bi-level	17329RMB	0.4042	7558.7kW	0.3792RMB/kWh

number of EVs is large, the total charging power is larger than the base load in Fig. 8. This causes the increase of the peak-valley gap in these control policies. From Table III, it can be found that the greedy charging policy performs the worst. The bilevel charging policy performs well both on the charging cost optimization and the matching degree optimization, while the delayed charging policy only performs well on the charging cost optimization. Due to this reason, the total spilled wind energy of the bilevel charging policy is 1581.5 kWh compared with 2807.2 kWh in the delayed charging policy.

VI. CONCLUSION

In this paper, a bilevel and multi-timescale scheduling problem is considered to match the EV charging load with the wind power. The bilevel MDP is used to formulate this stochastic programming problem. We propose a bilevel SBPI method to approximately solve this problem. The cost improvement property of this method is also proved. Numerical results demonstrate that our method can improve the wind power utilization and decrease the charging cost for a large number of EVs.

APPENDIX A
PROOF OF THEOREM 1

Proof: We prove it by induction. At stage M, there is $n = T_U(M-1)/T_L + 1$ and

$$J^U(\mu, X_M^u) = \sum_i q(S_n^i|X_M^u)\mathbf{E}\left[\sum_{\tau=n}^{N} C(S_\tau^i, \pi(S_\tau^i))|S_n^i\right] = \sum_i q(S_n^i|X_M^u) J^L(\pi, S_n^i)$$

where $q(S_n^i|X_M^u)$ denotes the probability that the lower level state is S_n^i when the upper level state is X_M^u. We have the following relationship:

$$J^U(\mu, X_M^u) - J^L(\pi, S_n^l) \le (J^L(\pi, S_n^l) + \epsilon_M) - J^L(\pi, S_n^l) = \epsilon_M$$
$$J^U(\mu, X_M^u) - J^L(\pi, S_n^l) \ge (J^L(\pi, S_n^l) - \epsilon_M) - J^L(\pi, S_n^l) = -\epsilon_M.$$

Then, there is $|J^U(\mu, X_M^u) - J^L(\pi, S_n^l)| \le \epsilon_M$.

Suppose at stage $m+1$, there is $|J^U(\mu, X_{m+1}^u) - J^L(\pi, S_{n+\alpha}^l)| \le \sum_{\tau=m+1}^{M} \epsilon_\tau$, then at stage m, there is

$$J^U(\mu, X_m^u) = \sum_i q(S_n^i|X_m^u)\left\{\mathbf{E}\left[\sum_{\tau=n}^{n+\alpha-1} C(S_\tau^i, \pi(S_\tau^i))|S_n^i\right] + \sum_j P(S_{n+\alpha}^j|S_n^i, \pi(S_n^i)) J^U(\mu, X_{m+1}^u)\right\}$$

where $\alpha = T_U/T_L$ and $P(S_{n+\alpha}^j|S_n^i, \pi(S_n^i))$ denote the transition probability that the lower level state at stage $n+\alpha$ is $S_{n+\alpha}^j$ when the lower level state at stage n is S_n^i and action $\pi(S_n^i)$ is chosen. Based on the assumption at stage $m+1$, there is

$$\begin{aligned} J^U(\mu, X_m^u) &\le \sum_i q(S_n^i|X_m^u)\left\{\mathbf{E}\left[\sum_{\tau=n}^{n+\alpha-1} C(S_\tau^i, \pi(S_\tau^i))|S_n^i\right] + \sum_j P(S_{n+\alpha}^j|S_n^i, \pi(S_n^i))\left(J^L(\pi, S_{n+\alpha}^j) + \sum_{\tau=m+1}^{M}\epsilon_\tau\right)\right\} \\ &= \sum_i q(S_n^i|X_m^u) J^L(\pi, S_n^i) + \sum_{\tau=m+1}^{M}\epsilon_\tau \\ &\le J^L(\pi, S_n^l) + \epsilon_m + \sum_{\tau=m+1}^{M}\epsilon_\tau \le J^L(\pi, S_n^l) + \sum_{\tau=m}^{M}\epsilon_\tau. \end{aligned}$$

Similarly, there is $J^U(\mu, X_m^u) \ge J^L(\pi, S_n^l) - \sum_{\tau=m}^{M}\epsilon_\tau$. Then, we can have

$$|J^U(\mu, X_m^u) - J^L(\pi, S_n^l)| \le \sum_{\tau=m}^{M}\epsilon_\tau \quad \forall S_n^l \in X_m^u.$$

□

APPENDIX B
PROOF OF THEOREM 2

Proof: It is proved in [22] that $\delta_m \ge 0$. By using Theorem 1, there is the following relationship:

$$\begin{aligned} J^L(\pi^{b'}, S_n^l) &\le J^U(\mu^I, X_m^u) + \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} \\ &= J^U(\mu^b, X_m^u) + \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} - \delta_m \\ &\le J^L(\pi^b, S_n^l) + \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{LU}} + \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} - \delta_m. \end{aligned} \quad (36)$$

Similarly, there is

$$\begin{aligned} J^L(\pi^{b'}, S_n^l) &\ge J^U(\mu^I, X_m^u) - \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} \\ &= J^U(\mu^b, X_m^u) - \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} - \delta_m \\ &\ge J^L(\pi^b, S_n^l) - \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{LU}} - \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}} - \delta_m. \end{aligned} \quad (37)$$

By combining inequations (36) and (37), (33) can be obtained. When $\delta_m \ge \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{LU}} + \sum_{\tau=m}^{M}\epsilon_\tau^{\mathrm{UL}}$, we can obtain $J^L(\pi^{b'}, S_n^l) \le J^L(\pi^b, S_n^l)$ from (36). □

REFERENCES

[1] *2014 Annual Review on Wind Power Installation in China*. Chin. Wind Energy Assoc., Beijing, China, 2015.

[2] National Energy Administration of China. *The Development Plan for the Industry of Energy Saving and Renewable Energy Vehicles (2012-2020) by the State Council*, accessed on Nov. 30, 2015. [Online]. Available: http://www.nea.gov.cn/2012-07/10/c_131705726.htm

[3] J. Nair, S. Adlakha, and A. Wierman, "Energy procurement strategies in the presence of intermittent sources," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 85–97, Jun. 2014.

[4] J. Seixas, S. Simões, L. Dias, A. Kanudia, P. Fortes, and M. Gargiulo, "Assessing the cost-effectiveness of electric vehicles in European countries using integrated modeling," *Energy Policy*, vol. 80, pp. 165–176, May 2015.

[5] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 1, pp. 67–78, Jan. 2013.

[6] M. D. Galus, S. Koch, and G. Andersson, "Provision of load frequency control by PHEVs, controllable loads, and a cogeneration unit," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4568–4582, Oct. 2011.

[7] N. Liu *et al.*, "A heuristic operation strategy for commercial building microgrids containing EVs and PV system," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2560–2570, Apr. 2015.

[8] M. H. K. Tushar, C. Assi, M. Maier, and M. F. Uddin, "Smart microgrids: Optimal joint scheduling for electric vehicles and home appliances," *IEEE Trans. Smart Grid.*, vol. 5, no. 1, pp. 239–250, Jan. 2014.

[9] K. Mets, F. De Turck, and C. Develder, "Distributed smart charging of electric vehicles for balancing wind energy," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2012, pp. 133–138.

[10] Q. Huang, Q.-S. Jia, L. Xia, X. Guan, and X. Xie, "EV charging load scheduling following uncertain renewable energy supply by stochastic matching," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2014, pp. 137–142.

[11] T. Zhang, W. Chen, Z. Han, and Z. Cao, "Charging scheduling of electric vehicles with local renewable energy under uncertain electric vehicle arrival and grid power price," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2600–2612, Jul. 2014.

[12] W. Yao, J. Zhao, F. Wen, Y. Xue, and G. Ledwich, "A hierarchical decomposition approach for coordinated dispatch of plug-in electric vehicles," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2768–2778, Aug. 2013.

[13] J. A. P. Lopes, F. J. Soares, and P. M. R. Almeida, "Integration of electric vehicles in the electric power system," *Proc. IEEE*, vol. 99, no. 1, pp. 168–183, Jan. 2011.

[14] B.-G. Kim, S. Ren, M. van der Schaar, and J.-W. Lee, "Bidirectional energy trading and residential load scheduling with electric vehicles in the smart grid," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1219–1234, Jul. 2013.

[15] M. Carrión and R. Zárate-Miñano, "Operation of renewable-dominated power systems with a significant penetration of plug-in electric vehicles," *Energy*, vol. 90, pp. 827–835, Oct. 2015.

[16] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1319–1327, Aug. 2008.

[17] Q. Huang, Q.-S. Jia, and X. Guan, "Multi-timescale optimization between distributed wind generators and electric vehicles in microgrid," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2015, pp. 671–676.

[18] J. W. Zack, "Overview of the current status and future prospects of wind power production forecasting for the ERCOT system," in *Proc. Wind Workshop III ERCOT Workshop*, Jun. 2009.

[19] S. Vandael, B. Claessens, M. Hommelberg, T. Holvoet, and G. Deconinck, "A scalable three-step approach for demand side management of plug-in hybrid vehicles," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 720–728, Jun. 26, 2013.

[20] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 1994.

[21] Q. Huang, Q.-S. Jia, and X. Guan, "Robust scheduling of EV charging load with uncertain wind power integration," *IEEE Trans. Smart Grid*, to appear.

[22] D. P. Bertsekas, J. N. Tsitsiklis, and C. Wu, "Rollout algorithms for combinatorial optimization," *J. Heuristics*, vol. 3, no. 3, pp. 245–262, 1997.

[23] Q. Huang, Q. S. Jia, Z. Qiu, X. Guan, and G. Deconinck, "Matching EV charging load with uncertain wind: A simulation-based policy improvement approach," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1425–1433, May 2015.

[24] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 2. Belmont, MA, USA: Athena Scientific, 2012.

[25] D. P. Bertsekas, "Rollout algorithms: An overview," in *Proc. 38th IEEE Conf. Decision Control*, Phoenix, AZ, USA, Dec. 7–10, 1999, pp. 448–449.

[26] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA, USA: Athena Scientific, 1996.

[27] *Weather Station at Tsinghua University, China*, accessed on Nov. 30, 2015. [Online]. Available: http://climate.dest.com.cn/

[28] *Vestas V100-2.0 MW*, accessed on Nov. 30, 2015. [Online]. Available: http://vestas.com

[29] I. M. Giammanco, J. L. Schroeder, and M. D. Powell, "Observed characteristics of tropical cyclone vertical wind profiles," *Wind Struct.*, vol. 15, no. 1, pp. 65–86, 2012.

[30] *2013 Beijing Transport Annual Report*, Beijing Transp. Res. Center, Beijing, China, 2013.

[31] *BYD e6*, accessed on Nov. 30, 2015. [Online]. Available: http://www.byd.com/na/e6/e6.html

Qilong Huang (S'14) received the B.E. degree in automation from Xi'an Jiaotong University, Xi'an, China, in 2010. He is currently pursuing the Ph.D. degree with the Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing, China.

His current research interests include energy management of smart grid, wind power integration, and optimization of large-scale systems.

Qing-Shan Jia (S'02–M'06–SM'11) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2002 and 2006, respectively.

He was a Visiting Scholar at Harvard University, Cambridge, MA, USA, in 2006, and the Hong Kong University of Science and Technology, Hong Kong, in 2010. He is currently an Associate Professor with the Center for Intelligent and Networked Systems, Department of Automation, TNLIST, Tsinghua University. His current research interests include theories and applications of discrete event dynamic systems and simulation-based performance evaluation and optimization of complex systems.

Xiaohong Guan (M'93–SM'95–F'07) received the B.S. and M.S. degrees in control engineering from Tsinghua University, Beijing, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, CT, USA, in 1993.

He was a Senior Consulting Engineer with PG&E, San Francisco, CA, USA, from 1993 to 1995. He visited the Division of Engineering and Applied Science, Harvard University, Cambridge, MA, USA, from 1999 to 2000. Since 1995, he has been with the Systems Engineering Institute, Xi'an Jiaotong University, Xi'an, China, where he has been appointed Cheung Kong Professor of Systems Engineering since 1999, and Dean of the School of Electronic and Information Engineering since 2008. Since 2001, he has also been the Director of the Center for Intelligent and Networked Systems, Tsinghua University, and served as Head of the Department of Automation from 2003 to 2008. His current research interests include complex networked systems, including smart power grids, planning and scheduling of electrical power and manufacturing systems, and electric power markets.