Matching Uncertain Renewable Supply with Electric Vehicle Charging Demand-A Bi-Level Event-Based Optimization Method

Teng Long  
Center For Intelligent and Networked Systems, Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

Qing-Shan Jia  
Center For Intelligent and Networked Systems, Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

Follow this and additional works at: https://dc.tsinghuajournals.com/complex-system-modeling-and-simulation

Recommended Citation

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in Complex System Modeling and Simulation by an authorized editor of Tsinghua University Press: Journals Publishing.
Matching Uncertain Renewable Supply with Electric Vehicle Charging Demand—A Bi-Level Event-Based Optimization Method

Teng Long and Qing-Shan Jia

Abstract: The matching between dynamic supply of renewable power generation and flexible charging demand of the Electric Vehicles (EVs) can not only increase the penetration of renewables but also reduce the load to the state electric power grid. The challenges herein are the curse of dimensionality, the multiple decision making stages involved, and the uncertainty of both the supply and demand sides. Event-Based Optimization (EBO) provides a new way to solve large-scale Markov decision process. Considering different spatial scales, we develop a bi-level EBO model in this paper which can both catch the changes on the macro and micro levels. By proper definition, the size of event space stays fixed with the scale of the problem, which shows good scalability in online optimization. Then a bi-level Q-learning method is developed to solve the problem iteratively. We demonstrate the performance of the method by numerical examples. Our method outperforms other methods both in performance and scalability.

Key words: electric vehicle; event-based optimization; cyber physical energy system

1 Introduction

The growing population of Electric Vehicles (EVs) in recent years has made controlled EV charging a popular research topic[1]. Although the stochastic matching of renewable power generation and EV charging demand could reduce the CO₂ emission and save the overall charging cost, the uncontrolled charging behavior will affect the stability of the state power grid[2]. Many efforts were devoted to promote the application of local wind turbines of skyscrapers[3]. Yang et al.[4] showed the potential of using on-site wind energy generations of skyscrapers to supply EVs charging in the smart grid. However, it may be unrealistic and rude to control privately owned EVs directly. Thus, we consider the controlling of EV fleets operated by a company like Uber, Car2go, and so on in this work since they are interested to reduce the operating cost by arranging the EV charging schedule under renewable energy supply.

To solve the large-scale EV charging scheduling problem is challenging due to the curse of dimensionality. With the increase of the number of EVs, both the state and action space grow exponentially. Meanwhile, it contains a multiple decision-making stages since the charging action now will affect the future states. Furthermore, both the wind energy generation and the charging loads of EV fleets are uncertain and sequentially revealed. There exist many efforts to solve the charging scheduling problem. We provide a brief review in Section 2. Despite these progresses, the current work usually falls short in scalability and solving speed.

Compared with the existing works on the similar topics, this paper considers the charging scheduling problem and makes the following main contributions. First, we develop an innovative bi-level Event-Based Optimization (EBO) model. In the upper level, the central controller decides how to dispatch the power that combines the wind power and the utility grid. In the lower level, the charging for each EV is scheduled. By proper definition, the number of events can be constant, so as to reduce the huge state space in the original Markov Decision Process (MDP) model. Second, we develop a bi-level Q-learning algorithm to solve this
bi-level EBO. It is an online algorithm which can be updated in real-time market. Third, we demonstrate the effectiveness of the proposed method via a series of numerical experiences. Our method outperforms other methods in performance and shows advantages in scalability. The impacts of wind power supply and event parameters are assessed.

The remainder of this paper is organized as follows. We give the related literature review in Section 2; formulate the original problem as an MDP in Section 3; introduce the main results in Section 4, including the bi-level EBO model and the bi-level Q-learning algorithm; conduct the numerical experiments in Section 5; and briefly conclude in Section 6.

2 Literature Review

The charging control of EVs have been active areas of research for years. In light of this, those works are different from objective functions, such as peak procurement minimization\(^\text{[5]}\), battery health protection\(^\text{[6]}\), state frequency stabilization\(^\text{[7]}\), and valley filling\(^\text{[8]}\). Various approaches are proposed, such as gradient projection based method\(^\text{[9]}\), branch-and-bound algorithms\(^\text{[10]}\), and benders decomposition\(^\text{[11]}\).

An off-line algorithm could be applied to obtain the optimal charging policy if the future information could be known in advance\(^\text{[12]}\). When taking the uncertainties under consideration in the real-time market, Model Predictive Control (MPC)\(^\text{[13]}\) and robust control\(^\text{[14]}\) are applied. Moghaddass et al.\(^\text{[15]}\) proposed a mixed-integer multi-objective optimization algorithm where two scheduling frameworks (static and dynamic) were considered. Wang et al.\(^\text{[16]}\) considered the EV admission control, pricing, and charging scheduling together and introduced a joint optimization method to improve the profit of the charging station. In addition to centralized control, a distributed algorithm has been developed in Ref. \(^\text{[4]}\) from the perspective of building managers. Similarly, Liu et al.\(^\text{[8]}\) introduced a shrunken-primal-dual sub-gradient method which can be applied in the decentralized optimization scheme. Considering the Vehicle-To-Grid (V2G) technology, Sun et al.\(^\text{[17]}\) used the EVs as moving storage systems to be able to charge/discharge while parking. Detailed review of related datasets and optimization methods can be found in Ref. \(^\text{[18]}\). Among these works, the idea of EV Aggregators (EVAs) is an effective way to schedule EVs fleet which can save computation and communication burden\(^\text{[19]}\).

Among the various optimization models, the MDP provides the most detailed system dynamics. However, it usually has the difficulty of the curse of dimensionality. EBO shows great potential to reduce the state and action space. An event \(e\) is defined as a set of state transitions that have certain common properties\(^\text{[20]}\) and EBO is mathematically equivalent to the partially observable MDP\(^\text{[21]}\). Generally, the event space is much smaller compared with the original MDP and many works applied EBO to address the EV charging problem\(^\text{[18, 22]}\). Note that these works have not explored the fact that some events only capture local state changes, while some events capture global state changes. Besides, the scheduling objects of those work mentioned above are mainly privately owned EVs which make privacy security a worrying issue, and little literature has considered EV fleets managed by the commercial company. Different from the existing papers, this paper focuses on scheduling the charging schedule of EVs managed by the commercial company with on-site wind power to minimize the operation cost.

Our research group has carried out some researches on this scheduling problem\(^\text{[23, 24]}\), but some core difficulties of large-scale EV charging (state space and action space explosion) are still not completely solved, which limits the scalability of proposed algorithms. In this work, we innovatively use incomplete Beta function in the lower level to reduce the policy state to \(O(2)\). Based on that, we propose a bi-level EBO model which distinguishes global and local events on different spatial scales. With proper definition, the number of events stays as a constant in our model which greatly enhances the scalability of our algorithm.

3 Problem Formulation

The main stakeholder in the problem description is an Independent System Operator (ISO, representing the commercial company), who manages \(K\) EVAs and \(N\) EVs. The optimization goal is to minimize the operating cost. The framework of the whole system is shown in Fig. 1. Charging stations of one region is managed by an EVA, who has locally wind power generators on the high-rise buildings and a certain number of charging piles. We only utilize the on-site wind power and do not consider energy storage. When an EV \(i, i \in \{1, \ldots, N\}\) is parking at a charging station in the region \(k, k \in \{1, \ldots, K\}\), the information of this EV will be delivered to the EVA \(k\). Note that an EV may travel among different regions and report to different
EVs. In other words, the affiliation of an EV with an EVA is dynamic and depends on the geographic location of the EV when it parks. This is different from many existing works where a fleet of EVs are steadily aggregated by the company. EVAs update the information including the EV states and renewable energy prediction to the ISO.

We made the following assumptions in the formulation unless clearly stated otherwise.

1. The charging power \( P \) of charging piles is constant.
2. The wind power is free.
3. EVs will update their information (such as the length of the parking and the required State-Of-Charge (SOC) by the end of the parking) to EVAs.

The Assumption 1 makes sense since a constant charging power may prolong the lifetime of the battery. Assumptions 2 and 3 are reasonable because the wind power generators and EVs are assets of the company. Meanwhile, when the renewable energy is insufficient for the EV charging, the utility grid may be utilized at a price.

Without loss of generality, we consider a discrete-time decision-making process. Let \( s_t = (W_t, L_{t,i}, E_{t,i}, G_{t,i}) \in S, i \in \{1, \ldots, N\} \) be the system state, where \( t \) is the time index and takes nonnegative integers (i.e., from the set \( \mathbb{Z}^+ \)). \( W_t \) is the wind power generation. \( L_{t,i} \) denotes the remaining parking time of EV \( i \) and \( E_{t,i} \) is the remaining charging demand. \( G_{t,i} \) denotes the geographical state of EV \( i \), where \( G_{t,i} = k \) means EV \( i \) is parking at EVA \( k \) and \( G_{t,i} = 0 \) means EV \( i \) is on the road. We calculate the wind power generation at time \( t \) as follows:

\[
W_{t,k} = \begin{cases} W_{\text{cap}}, & v^\text{in} \leq v_t \leq v^\text{out}; \\ W_{\text{cap}} \left( \frac{v_{t,k}}{v_{\text{cut-in}}} \right)^3, & v^\text{in} \leq v_t \leq v^\text{r}; \\ 0, & \text{otherwise} \end{cases} \tag{1}
\]

where \( v^\text{in}, v^r, \) and \( v^\text{out} \) are the cut-in speed, rated speed, and cut-out speed, respectively. \( W_{\text{cap}} \) is the capacity of the wind turbine and \( v_{t,k} \) is the wind speed of region \( k \). Assuming that the wind power is used preferentially by the current EVA and can be shared by other EVAs. To ensure the charging requests to be fulfilled before leaving, we have the following constraint:

\[
E_{t,i} \geq L_{t,i} P \geq E_{t,i} \tag{3}
\]

where \( E_{\text{cap}} \) is the battery capacity of EVs. Formula (3) means that the remaining parking time should be sufficient to satisfy the EV charging request. When the division of the EV charging demand \( E_{t,i} \) by \( P \) is not an integer, we approximate that by the minimal integer that is not smaller than the division.

Let \( a_t \in \mathcal{B}^N \), where \( \mathcal{B} = \{0, 1\} \), be the charging action at time \( t \), where \( a_t(i) = 1 \) (or 0) denotes that the \( i \)-th EV is scheduled to be (or not to be) charged by the ISO at time \( t \). Let \( C_t(s_t, a_t) \) be the one-step cost function. We have

\[
C_t(s_t, a_t) = \zeta_t \max \left\{ P \sum_{i=1}^{N} a_t(i) - W_t, 0 \right\} \tag{4}
\]

where \( \zeta_t \) denotes the Time-of-Usage (ToU) price of
the utility grid, and \(\max\{P \sum_{i=1}^{N} a_t(i) - W_t, 0\}\) is the amount of power bought from the utility grid. When the wind power is sufficient, the purchase cost will be zero. Consider the \(M\)-step long-term total cost,

\[
J(\mu) = \left\{ \sum_{t=0}^{t_0+M} C_t(s_t, \mu(s_t)) | s_{t_0} \right\}
\]

where \(\mu : S \rightarrow A\) is a charging policy that decides the charging of all EVs based on the states. Now we have formulated the original MDP model,

\[
\min_{\mu} J(\mu),
\]

s.t. All EV charging demand can be satisfied

Denote this problem as P1. As mentioned in Section 1, this problem is difficult to solve since its state and action spaces increase exponentially fast with regard to the problem scale \(N\). In the next section, we will present a bi-level EBO model which significantly reduces the problem scale and accelerates the solving process.

4 Main Result

The problem P1 is computationally inefficient. We develop the bi-level EBO formulation to reduce the state space and develop the bi-level event-based \(Q\)-learning.

4.1 Bi-level EBO model

The EV charging scheduling can be divided into two steps, how many EVs will be charged in each EVA and which EVs will be scheduled to be charged. We propose a bi-level model to solve those two problems, respectively. In the upper level, the ISO decides the total amount of power to charge in the EVA \(k\), denoted as \(P_{t,k}\). In the lower level, based on the power dispatched from the upper level, each EVA decides whether each EV parked in the region can be charged. In order to capture the system dynamics at a different scale, we will define macro events and micro events for the upper- and lower-level EBO, respectively.

4.1.1 Upper-level EBO

The number of EVs that need to be charged \(N_{t,k}^{c}\) and the number of EVs that must be charged \(N_{t,k}^{m}\) are the information that reflect the macro changes in EVAs. Thus, the EVA will update those information to the ISO to support the decision making. The definitions of \(N_{t,k}^{m}\) and \(N_{t,k}^{c}\) are as follows:

\[
N_{t,k}^{m} = \sum_{i=1}^{N} I(E_{t,i} = PL_{t,i} \text{ and } G_{t,i} = k) \tag{7}
\]

\[
N_{t,k}^{c} = \sum_{i=1}^{N} I(E_{t,i} > 0 \text{ and } G_{t,i} = k) \tag{8}
\]

\[
N_{t,k}^{c} \geq N_{t,k}^{m} \tag{9}
\]

where \(I(\cdot)\) is the indicator function.

(1) Macro event: In metropolitan cities, it is not scalable for the ISO to track the states of each EV. Instead, the EVAs may report some statistics of the states of EVs in the region, which can help the ISO to decide the charging schedule for EVs. \(W_t\), \(N_{t,k}^{c}\), and \(N_{t,k}^{m}\) are such statistics. We introduce a mismatch index \(M_t\) between the renewable energy and charging requests, which is defined as follows:

\[
M_t(s_t) = \begin{cases} 
B_2, & B_1 > 0; \\
0, & B_1 = 0 
\end{cases} \tag{10}
\]

\[
B_1 = \sum_{k=1}^{K} N_{t,k}^{c} \tag{11}
\]

\[
B_2 = \max\left\{PB_1 - \max\left\{W_t - P \sum_{k=1}^{K} N_{t,k}^{m}, 0\right\}, 0\right\} \tag{12}
\]

where \(B_2\) is the remaining schedulable loads after consuming all the wind power at time \(t\) and \(B_1\) denotes the total number of schedulable EVs. Note that when \(M_t\) takes larger/smaller values, this means that there are more/less flexibilities in the charging schedule. We use macro events to describe different levels of flexibility, i.e.,

\[
e_{t}^{m}(j) \equiv \{(s, s') | M_t(s) \in \left(\frac{j - 1}{R}, \frac{j}{R}\right]\} \tag{13}
\]

where \(j = 1, \ldots, R\) and the flexibility is divided into \(R + 1\) levels. \(s\) and \(s'\) denote the states at time \(t\) and \(t + 1\), respectively. Denote the macro events at time \(t\) as \(E_{t}^{m} \in \mathcal{E}_{t}^{m} = \{e_{t}^{m}(j)\}, j = 0, 1, \ldots, R\). Define \(e_{t}^{m}(0)\) as the event that the renewable output can support all the charging requests of EVs.

(2) Macro action: As mentioned above, the ISO makes the decision of energy purchasing and dispatching, namely the total power bought from the utility grid \(P_{t}^{B}\)
and the dispatched power for each EVA, i.e., \( P_{t,k} = P_{t,k}^{m} \). This may be done following two steps. First, the power \( P_{t,k}^{m} \) for EVs which must be charged should be satisfied, i.e., \( P_{t,k}^{m} \leq P_{t,k}^{m} \).

Second, for \( N_{t,k}^{c} \) EVs that can be scheduled in the EVA \( k \), the ISO needs to decide how many of them will be charged at this time step. Since the lower-level EVAs should be treated equally to address the same level of charging needs, we denote a charging ratio \( \alpha_{t} \) for all EVAs, and \( \alpha_{t}(k)N_{t,k}^{c} \) EVs that are schedulable can be charged now. Thus, the actual power received by the EVA \( k \) is

\[
P_{t,k} = P_{t,k}^{m} + P\alpha_{t}N_{t,k}^{c} \tag{14}
\]

Thus, we have

\[
P_{t}^{B} = \sum_{k=1}^{K} P_{t,k} - W_{t} \tag{15}
\]

Equation (15) ensures that the charging power (combining the renewable energy and the electricity) is sufficient for the set of EVs that must be charged, i.e.,

\[
P \sum_{k=1}^{K} N_{t,k}^{c} \geq P_{t}^{B} + W_{t} \geq P \sum_{k=1}^{K} N_{t,k}^{m} \tag{16}
\]

Now we can denote the aforementioned macro action as

\[
A_{t}^{m} = (P_{t}^{B}, \alpha_{t}) \tag{17}
\]

(3) Constraints: The following constraints should be satisfied in the upper level,

\[
0 \leq N_{t,k}^{full} \leq N_{t,k}^{c} \tag{18}
\]

\[
N_{t,k}^{m} \leq N_{t,k}^{c} \tag{19}
\]

where \( N_{t,k}^{full} \) denotes the number of EVs fully charged at EVA \( k \) at time step \( t \). The constraint Formula (18) guarantees that the number of EVs that need to be charged is always larger than the number of EVs that is already fully charged. The constraint Formula (19) means that the upper bound of the number of EVs that must be charged is the number of EVs that need to be charged.

(4) System dynamics: We only consider dynamics that affects macro-level statistics in the upper level, which are

\[
N_{t+1,k} = N_{t,k} + N_{t,k}^{m} - N_{t,k}^{out} \tag{20}
\]

\[
N_{t+1,k}^{c} = N_{t,k}^{c} + N_{t,k}^{in} - N_{t,k}^{full} \tag{21}
\]

\[
N_{t+1}^{c} = N - \sum_{k=1}^{K} N_{t+1,k} \tag{22}
\]

where \( N_{t,k}, N_{t,k}^{in}, \) and \( N_{t,k}^{out} \) denote the number of EVs parking, arriving, and leaving, respectively. \( N_{t,k}^{c} \) is the number of EVs on the roads.

4.1.2 Lower-level EBO

We consider from the perspective of EVA in the lower level and give the model for each EVA identically. The EVA need to arrange the local charging load of EVs, given the total power \( P_{t,k}^{B} \) from the upper level. Define the urgency index function for EV \( i \) as follows:

\[
\Delta_{t}(i) = \frac{E_{t,i}}{L_{t,i}^\frac{1}{\beta}} \tag{23}
\]

which describes the urgency of charging. For all EVs parking in the EVA \( k \), i.e., \( j \in \{i | G_{t,j,\beta} = k\} \), we may rank and count EVs according to their urgency index, from small to large, and may obtain a monotonically nondecreasing curve, which is called the Ordered Urgency Curve (OUC). We normalize the vertical and the horizontal ranges of the OUC to \([0, 1]\), which is then referred to as the Standardized OUC (SOUC). Some examples of SOUCs are shown in Fig. 2.

Note that these kinds of SOUCs can be described by the incomplete Beta function\(^{[29]}\), that is,

\[
F(y|\alpha, \beta) = \int_{0}^{y} f(z|\alpha, \beta) dz \tag{24}
\]

where the beta density \( f(z|\alpha, \beta) \), \( \alpha > 0, \beta > 0 \) is defined as

\[
f(z|\alpha, \beta) = C y^{\alpha-1}(1-z)^{\beta-1} \tag{25}
\]

where \( C = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \) and \( \Gamma(\cdot) \) is the gamma function. Then the SOUC is determined by the inverse function of \( F(y|\alpha, \beta) \).

\[
\Delta_{t}(x|\alpha, \beta) = F^{-1}(x|\alpha, \beta) = F(x|1/\alpha, 1/\beta) \tag{26}
\]

where \( F(\cdot, \cdot) \) can be evaluated via numerical approximation formulas. In this way, we can use two parameters \( \alpha \) and \( \beta \) to represent the SOUC of an EVA. According to Ref. [30], \( \log \alpha \) and \( \log \beta \) chosen from \((-2.5, 2.5)\) sufficiently represent the five types of SOUCs considered. We set the level of discreteness on one dimension as \( V \) which means there will be \( V^2 \) pairs of \((\log \alpha, \log \beta)\).

![Fig. 2 Five typical SOUCs.](image-url)
(1) **Micro event:** We now define the micro-level events to capture this important information in the lower level, which is

\[ e^j(t) = \{(s, s')|SOUC(s) \text{ is best approximated} \] 

by the \( j \)-th pair of\((\log \alpha, \log \beta)\),

\[ j = 1, \ldots, V^2 \] (27)

Denote the micro events at time \( t \) as \( E^I_t \in \mathcal{E}^I = \{e^j(t)\}, j = 0, 1, \ldots, V^2 \) which describe the emergency situations of each EVA.

(2) **Micro action:** The EVA \( k \) need to arrange the charging process for each EV in the region \( k \). First, the EVA charges EVs that must be charged immediately. The amount of power consumed in this step is \( P_{N_k} \). Second, the EVA distributes the remaining power \( P_{t+1, k} = P_{N_k} \) among the rest EVs which still has flexibility. We define two actions \( A_j = \{A_1, A_2\} \) to map the remaining power with the EVs in this work, which are shown as follows,

- \( A_1 \): A deterministic action. Distribute the power according to the SOUC.
- \( A_2 \): A probabilistic action. Randomly pick EVs to charge according to a probability that is proportional to the SOUC.

With a slight abuse of notation, \( A_1 \) and \( A_2 \) can be written in the form of

\[ Z^I_t = [z^I_1(t), z^I_2(t), \ldots, z^I(N)] \] (28)

where \( z^I_i(t) = 1 \) (or 0) means EV \( i \) is (not) scheduled to be charged.

(3) **Constraints:** We list the constraints of EVAs as follows:

\[ L_{t,i}P \geq E_{t,i} \geq 0 \] (29)

\[ E^{\text{cap}} \geq E_{t,i} \geq 0 \] (30)

The constraint Formula (29) guarantees that the EV charging needs can be completed in time. The constraint Formula (30) shows that the remaining charging demand should be lower than the battery capacity \( E^{\text{cap}} \).

(4) **System dynamics:** Define

\[ U_{t,i} = I(G_{t,i} > 0) \] (31)

where \( U_{t,i} \) denotes whether the EV \( i \) is parking, then may be charged (or not). After the charging decisions have been made in the lower level, the state transition are defined as follows\(^{[31]}\):

\[ L_{t+1,i} = \begin{cases} L_{t,i} - \Delta T, & \text{if } U_{t,i} = 1; \\
\tau_{t+1,i}, & \text{if } U_{t,i} = 0 \text{ and } U_{t+1,i} = 1; \\
0, & \text{if } U_{t,i} = 0 \text{ and } U_{t+1,i} = 0 \end{cases} \] (32)

where \( \Delta T \) is the length of the time interval. \( \tau_{t+1,i} \) is the random newly parking time and \( \eta_{t+1,i} \) denotes the random charging request. Thus, we have the following constraint:

\[ \eta_{t+1,i} = \tau_{t+1,i}P \] (34)

### 4.1.3 Objective function

Although the electricity purchase only happens in the upper level (we just dispatch the power to EV As and then they will find a way to allocate it in the lower level), the macro and micro actions will determine the state transition of the system together, and affect the future events. We define the cost function at time \( t \) as

\[ C_t(E^u_t, E^l_t, A^u_t, A^l_t) = \max \left\{ \sum_{k=1}^K P_{t,k} - W_t, 0 \right\} \] (35)

and consider the \( M \)-step total cost as follows:

\[ J(\mu^u, \mu^l) = \] (36)

\[ \sum_{i=0}^{t_0+M} C_i(E^u_i, E^l_i, \mu^u(E^u_i), \mu^l(E^l_i))|E^u_{t_0}, E^l_{t_0} \]

where \( E^u_{t_0} \) and \( E^l_{t_0} \) are the initial events at time \( t_0 \); \( \mu^u \in \mathcal{E}^u \) and \( \mu^l \in \mathcal{L}^l \) are given event-based macro and micro policies, respectively.

Then the EBO model is to seek for an optimal charging policy \( \mu^u \) and \( \mu^l \) that can minimize the \( M \)-step total cost, that is,

\[ (\mu^u, \mu^l) = \arg \min_{\mu^u, \mu^l} J(\mu^u, \mu^l) \] (37)

Denote this problem as P2 which is an extended version of P1. But P2 explores the structural property of P1 to distinguish local events from global events and may lead to computational advantage, which is shown in Table 1.

In P2, the event and action spaces will be fixed after all the parameters are set. While in P1, the state and action spaces increase exponentially with \( N \).

### 4.2 Bi-level Q-learning algorithm

Generally speaking, the Q-learning has always been an effective method in solving EBO problems\(^{[32]}\). However,
the classical Q-learning method cannot be directly applicable to the problem with a hierarchical structure. We develop a bi-level event-based Q-learning algorithm to solve the problem P2 in this subsection. It is an online algorithm which can be updated in each step. We evaluate the Q-factors of each level, respectively. The Q-factor of the upper (or lower) level at time $t_0$ can be defined as follows:

$$Q^\mu_{t_0}(e, a) = E\{C_{t_0}(E_{t_0}, A_{t_0}) + \sum_{i=t_1}^{t_0+M} C_i^\mu(E_i)|E_{t_0} = e, A_{t_0} = a\}$$  \hspace{1cm} (38)

where $e$ and $a$ denote the macro (or micro) event and action. Equation (38) can also be written as an one-step iteration form,

$$Q^\mu_{t_k+1}(e, a) = Q^\mu_{t_k}(e, a) + (1/(N^\epsilon(e, a) + 1)) \cdot \left(C_{t_k}(e, a) + \lambda Q^\mu_{t_k}(e, a) - Q^\mu_{t_k}(e, \mu(e))\right)$$

where $N^\epsilon(e, a)$ is the number of times that the $(e, a)$ pair occurs and $\lambda$ is the learning rate. Then we update policies $\mu$ to $h^{[32]}$ at time $t_{k+1}$ which is defined by

$$h(e)_{k+1} := \text{arg max}_a \{Q^\mu_{t_k+1}(e, a)\}$$  \hspace{1cm} (40)

Considering the training of $T$ time steps, we summarize the algorithm in Algorithm 1.

Figure 3 shows the flowchart of Algorithm 1. As we can see, we do the optimization iteratively. In a time step, we first update policies $\mu^\mu$ to $\mu^\mu_k$ based on the macro event $e^\mu_k$, while the micro policy $\mu^\nu$ is fixed. Then in the lower level, we fix $\mu^\nu_k$ and update $\mu^\lambda$ instead based on the micro events. At the end of this time step, we can get the updated policies $\mu^\mu_k$ and $\mu^\nu_k$, which will be used in the next iteration.

5 Numerical Result

5.1 Parameter setting

In this section, a company, which has several wind turbines, EVAs, and EVs, is considered to evaluate the performance of proposed algorithm in terms of the cost reduction and scalability. To simulate the wind power generation, we follow three steps. First, the real wind speed at 12 m-height is collected at the Tsinghua University Weather Station. The SANY SE9320 wind turbine[33] is selected. Detailed parameter settings can be referred to Table 2.

Using Formula (1) and Eq. (2), we can convert the statistic wind speed data into the mean value of wind power generation which can be seen from Fig. 4. Second, we assume that the real wind energy follows a normal distribution with the standard deviation being 10% of the mean value. Third, we define the intensity index as follows:

$$\gamma = \frac{W_{\text{max}}}{PN}$$  \hspace{1cm} (41)

This parameter is used to simulate the wind power in different situations. Then the samples of uncertain wind power generation can be generated. The ToU price of electricity is also shown in Fig. 4.
Real driving data from Ref. [34] are utilized to generate the traffic flow and charging demand. Similarly, we also assume that the parking time follows the normal distribution with the standard deviation being 10% of the mean value. The charging demand is set to be proportional to the driving distance. The chi-square distribution is commonly used to fit the distribution of the driving distance [35]. The freedom degree of the chi-square distribution is set as 4.4868 [35]. The relationship between the charging demand $E^{eq}$ and the driving distance $d$ is as follows:

$$E^{eq} = [\omega d] P$$  \hfill (42)\]

where $\omega$ denotes the drive efficiency of electricity which we set $\omega = 0.195$ h/km. All the data we mentioned above are shown in Fig. 4. We use the BYD e6 as the EV model and the related parameter settings are shown in Table 2.

### 5.2 Performance analysis

In this subsection, a system with 12 EVs and 4 EV As is conducted to validate the effectiveness of the proposed methods. Thus, the following policies are tested:

1. Greedy policy. Charge EVs as soon as possible. This is the policy used in reality.
2. Delay policy. Standby and delay the charging as late as possible. In other words, only charge the EVs that must be charged.
3. Bi-level EBO (B-EBO) policy in Algorithm 1.
4. Multi-scale EBO (M-EBO) policy in Ref. [31].
5. BM-R: The optimal policy obtained by solving the original problem $P1$ with CPLEX and assuming that all the information is known in advance. This is the ideal policy which can not be applied to the real-time market and we will only use it as a measure of optimal performance.
6. BM-P: The policy obtained by solving the original problem $P1$ online with CPLEX and the 3-time-step prediction information.

To make a fair comparison, we set the price of the electricity in the upper and lower level be the same. We set $R = 10$, $V = 5$, and use 50 sample paths to evaluate the $Q$-factors. After a training process of $T = 21800$ time steps, we access the performance by setting $M = 24$ in the objective function to evaluate the total cost of a day. The numerical results of different policies in the case are summarized in Table 3. The ART denotes the abbreviation of average running time of one step policy iteration. The AUP represents the average unit price for charging.

In general, the optimal benchmark policy BM-R performs the best in reducing the total charging cost compared with other policies, though it is unrealistic since the information about the future cannot be known in advance. Meanwhile, the proposed B-EBO policy outperforms and improves the performance of the Greedy, Delay, M-EBO, and BM-P charging policies by 85.9%, 59.5%, 47.8%, and 18.2%, respectively. Figure 5 shows the detailed comparison about the charging cost and charged number of every hour. As Fig. 5b shows, the wind power is sufficient during 2:00–15:00 and turns low during 16:00–21:00. The Greedy and Delay policies schedule the EVs based on heuristic rules, thus ignoring
the impact of wind power and electricity price changes. The M-EBO, BM-P, and B-EBO are all trying to match the wind power generation. B-EBO policy does a better job. It wisely charges more EVs in advance when the wind energy is sufficient, thus avoiding charging EVs when the renewable energy is relatively weak.

Moreover, the decision (event) space of B-EBO is constant and equal to \((V^2 + R + 1)\), while the decision spaces of M-EBO, BM-R, and BM-P are \((3^N + R + 1)\), \(2^N T^P\), and \(2^N T\), respectively. \(T^P\) denotes the prediction steps. With the growth of EVs, the decision space of M-EBO, BM-R, and BM-P increases exponentially, which may take hours to solve the scheduling problem[36]. Thus, the M-EBO, BM-R, and BM-P methods are not suitable for solving a large-scale EV charging problem online. The proposed B-EBO method can solve the problem in a short time and iterate at every step in real-time market. We will analyze the scalability of the B-EBO method in the next subsection.

### 5.3 Scalability analysis

In the downtown of cities, the number of EVs of a company can be hundreds which puts forward high requirements for the scalability of the proposed scheduling algorithm. In this section, we increase EV number from 100 to 600. Meanwhile, the EVA number changes from 4 to 12. Since M-EBO, BM-R, and BM-P are not suitable for the large-scale scheduling, only Greedy, Delay, and B-EBO policies are compared in this section. Note that except for the number of EVs and EVAs, the parameters setting are the same as those described above.

Table 4 shows the total cost of different \(N\) and \(K\). The performance of B-EBO outperforms other policies in all cases. We fix \(K = 10\) and increase the EV number from \(N = 100\) to \(N = 800\) to assess the change of ART. The result is shown in Fig. 6. We can conclude that the

![Simulation results of every hour.](image)

**Fig. 5** Simulation results of every hour.

<table>
<thead>
<tr>
<th>(N)</th>
<th>(K)</th>
<th>Total cost (CNY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>4</td>
<td>854.86</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>1637.85</td>
</tr>
<tr>
<td>300</td>
<td>6</td>
<td>2445.62</td>
</tr>
<tr>
<td>400</td>
<td>8</td>
<td>3303.92</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>4099.02</td>
</tr>
<tr>
<td>600</td>
<td>12</td>
<td>4885.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(N)</th>
<th>(K)</th>
<th>Greedy</th>
<th>Delay</th>
<th>B-EBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>4</td>
<td>854.86</td>
<td>176.14</td>
<td>58.26</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>1637.85</td>
<td>322.66</td>
<td>140.98</td>
</tr>
<tr>
<td>300</td>
<td>6</td>
<td>2445.62</td>
<td>489.52</td>
<td>194.77</td>
</tr>
<tr>
<td>400</td>
<td>8</td>
<td>3303.92</td>
<td>636.57</td>
<td>289.98</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>4099.02</td>
<td>756.57</td>
<td>298.96</td>
</tr>
<tr>
<td>600</td>
<td>12</td>
<td>4885.74</td>
<td>891.00</td>
<td>303.52</td>
</tr>
</tbody>
</table>

![ART with different EV numbers.](image)

**Fig. 6** ART with different EV numbers.

ART shows a linear growth trend with the increase of system scale. It is reasonable since the size of event space is constant when the parameters are given, and the key factors affecting ART become the generation of simulation paths and the estimation of \(Q\)-factors which are not so sensitive to the scale of the system. Combining the experimental results above, the scalability of our algorithm has been proved.

### 5.4 Performance analysis under different wind power supplies

Different from EVs which have a relatively stable statistical distribution, the change of wind power supply is more dramatic. In this subsection, we analyze the
performance of B-EBO policy under different wind power supplies. We set \( N = 1000 \) and \( K = 20 \). We consider three cases with \( \gamma = 0.3, 0.8, \) and 1.3, and the other parameters setting are the same as those described above. The total cost for a day is shown in Table 5. The simulation results of every hour are shown in Fig. 7. The B-EBO policy outperforms other heuristic policies in all test scenarios with different wind power supplies. Comparing with Greedy and Delay policies, B-EBO policy saves more cost when the renewable energy generation is relatively sufficient (i.e., \( \gamma = 1.3, 0.8 \)). Moreover, even when \( \gamma = 0.3 \), B-EBO policy still reduces the cost of Greedy policy by 60\% and reduces the cost of Delay policy by 27\%. Note that B-EBO policy can better track the wind curve and adjust the policy accordingly, which is much better than both Greedy and Delay policies. It illustrates that the B-EBO policy has good robustness in the face of the uncertain wind power supply.

### 5.5 Parameter analysis

The macro and micro events in the bi-level EBO model are defined by discreteness. It is necessary to choose suitable parameters \( R \) and \( V \) considering the performance objective function. We compare the impact of different \( R \) and \( V \) in a system with 12 EVs and 4 EVAs when using the B-EBO policy. In addition, the parameter settings is the same as above.

The results can be seen from Table 6. In general, the best choice for \((R, V)\) pair is \((10, 5)\) and the maximum performance difference can be 66\% under different parameter settings which shows the importance to find proper parameters. On the dimension of \( V \), the total cost increases with the increase of values of \( V \). It is reasonable because more refined decentralizations will lead to the exponential growth of the policy space which may lead to only finding a sub-optimal policy. On the other hand, the total cost first decreases and then increases on the dimension of \( R \). It illustrates that rough decentralization means fewer macro events, which may also result in inaccurate descriptions of the system state and bad performance. A suitable parameter selection can balance the size of search space and the accuracy of the description so as to achieve better performance.

### 6 Conclusion

In this paper, we present a novel bi-level EBO model for the EV charging problem which can reduce the decision space significantly. A modified bi-level \( Q \)-learning algorithm is proposed to solve the bi-level EBO model. The performance, efficiency, and scalability of the proposed method are demonstrated by the numerical results. In this paper, EVs are only considered as portable loads that have uncertain charging demand. The
storage capacity of EVs will be considered to offset the uncertainty of wind power in future work.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (No. 2016YFB0901900), the National Natural Science Foundation of China (Nos. 62073182 and 61673229), and the 111 International Collaboration Project of China (No. BP2018006).

References


Qing-Shan Jia received the BS degree in automation and the PhD degree in control science and engineering from Tsinghua University, Beijing, China in 2002 and 2006, respectively. He was a visiting scholar at Harvard University in 2006, at Hong Kong University of Science and Technology in 2010, and at Massachusetts Institute of Technology in 2013. He is currently an associate professor at the Center for Intelligent and Networked Systems, Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University. His current research interests include theories and applications of cyber physical systems.

Teng Long received the BS degree in automation from Tsinghua University, Beijing, China in 2017. He is currently pursuing the PhD degree at the Center for Intelligent and Networked Systems, Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University. His current research interests include energy management of the smart grid, event-based optimization, and large-scale optimization problem.