Scheduling for Charging Plug-in Hybrid Electric Vehicles

Yunjian Xu and Feng Pan

Abstract—We construct a dynamic stochastic model to study the scheduling problem for battery charging of multiple (possibly a large number of) PHEVs. Our model incorporates the stochasticity in future PHEV arrival process and future renewable generation. The objective of scheduling is to maximize the overall social welfare, which is derived from total customer utility, the electricity cost associated with PHEV charging, and the non-completion penalty for not satisfying PHEVs’ charging requests. Through a dynamic programming formulation, we show the Less Laxity and Longer remaining Processing time (LLLP) principle: priority should be given to vehicles that have less laxity and longer remaining processing times, if the non-completion penalty (as a function of the additional time needed to fulfill the unsatisfied charging request) is convex. We introduce various forms of improved polices generated from a given heuristic policy according to the LLLP principle, and show that these improved polices can only improve social welfare, compared to the original heuristic.

I. INTRODUCTION

Building a sustainable future poses great challenges to many aspects of our current society. One of the major challenges is to create a sustainable environment by reducing greenhouse gas emissions. A potential technology to address the challenge is based on plug-in (hybrid) electric vehicles (PHEVs), which have been shown to produce less greenhouse gas emissions over their entire fuel cycle, compared to conventional vehicles [5], [14]. However, a large number of PHEVs can add considerable stress to an existing power grid, especially at the distribution network level [3], [10].

Large scale PHEV adoption has to be complemented with next generation power grids for which smart grid is usually used as an umbrella term. As a part of smart grids, renewable generation capacity is expanding rapidly to potentially reduce carbon dioxide footprint and dependence on fossil fuels. As non-dispatchable generation, renewable energy introduces variability into the energy portfolio, and further amplifies the difficulty of accommodating a large amount of loads from PHEVs. Demand-side management aims to reduce peak loads and to match variable renewable generation by reshaping electricity demands. Various topics related to demand-side management include load control with thermal mass in buildings [1], scheduling deferrable loads based on forecasted demand [18], and designing energy market for price elastic demands [7].

The focus of this paper is on demand-side management in the content of charging PHEVs. PHEV loads are usually deferrable and flexible to be scheduled. Proper management of the charging of PHEVs may help to stabilize the electric power system, as well as the retail price faced by customers. The scheduling of charging PHEVs receives much attention in recent years. To minimize the variance of PHEV load, a few recent papers propose several approaches based on mean value games [11], global optimization [6], and decentralized algorithms [8]. Although dynamic programming (DP)-based approaches have been employed to study the optimal control for power management of a single PHEV [16], [13], [12], there lacks a dynamic programming treatment on the scheduling problem of charging multiple PHEVs. On the electricity market side, the authors of [2] propose a decision support algorithm that enables demand side participation (such as PHEV loads) in wholesale electricity markets so as to offset the variations of renewable generation.

This work is also related to the literature on deadline scheduling problems. For a single processor scheduling problem, it is well known that simple scheduling algorithms, such as the earliest deadline first (EDF) policy [9] and the least-laxity first (LLF) policy [4], are optimal, in the sense that both EDF and LLF policies must be able to complete all tasks before their deadlines, if there exists some scheduling policy that can complete all tasks before their deadlines. However, this result does not apply to the scheduling problem studied in this paper, because (i) the cost associated with PHEV charging is stochastic (due to the uncertainty of future system load and future renewable generation), while the processor capacity in their model is fixed and constant, and (ii) the scheduling problem studied in this paper is for multiple (possibly a large number of) charging facilities that are able to charge PHEVs simultaneously, while in a single processor scheduling problem, the processor can work on only one task at a given time.1

In this paper, we consider a system with a finite number of PHEVs and an underlying power grid. A system operator (e.g., a scheduler of local parking garages or an operator of a local distributed network) schedules PHEV charging to maximize the overall social welfare, total customer utility (obtained from PHEV battery charging) minus the sum of the electricity cost associated with PHEV charging and the

1There is a literature on multiple-processor deadline scheduling problems; for a survey, see [17]. Most of previous works in the literature assume that each processor has constant processing capacity over the entire operation interval, while in our model, the capacity available for charging PHEV batteries can be time-variant, due to the volatility of system load and renewable generation.
penalty for unsatisfied charging requests. The operator’s
decision-making problem is essentially a stochastic deadline
scheduling problem with preemption and non-completion
penalty (for unsatisfied charging requests): every vehicle has
a departure time (deadline), and requires a certain amount
of processing time (e.g., the time needed to fully charge
its battery) before its deadline. We formulate the operator’s
decision-making problem as an infinite-horizon dynamic pro-
gramming (DP) problem. The constructed DP model includes
an exogenous source of uncertainty, the state of the power
grid, (e.g., representing the current system load and/or the
current renewable generation) that influences the supplier
cost associated with PHEV charging. It also incorporates the
stochasticity in future PHEV arrivals through an exogenous
state that evolves as a Markov chain.

The main result of this paper is to show the Less Laxity
and Longer remaining Processing time (LLLP) principle:
priority should be given to vehicles that have less laxity²
and longer remaining processing times, if the non-completion
penalty (as a function of the additional time needed to
complete the task) is convex, and if the operator does not
discount future cost³. For a given heuristic policy, we show
that an “updated” policy (cf. the formal definition prior to
Proposition 4.1) that follows the LLLP principle can only
improve the social welfare.

The rest of the paper is organized as follows. In Section II,
we introduce a system model for a finite number of PHEVs
and an underlying power grid. In Section III, we formulate
the PHEV scheduling problem as an infinite-horizon Markov
decision process. Main theoretical results, including the
LLLP principle, are introduced in Section IV. In Section
V, for a special case of our MDP (Markov decision process)
model where the power grid state represents the capacity
available for charging PHEV batteries, we compare the per-
formance of three heuristic policies. Numerical results show
that a stationary heuristic policy that cannot be updated by
the LLLP principle significantly reduces the time-averaged
cost, compared to the earliest deadline first (EDF) policy. In
Section VI, we make some brief concluding remarks.

II. MODEL

We consider an infinite-horizon dynamic model, which has
the following elements:

1) Consider an infinite-horizon discrete time model, where
time periods are indexed by \( t = 0, 1, \ldots \). 

2) We study the scheduling problem of \( N \) charging facili-
ties. Each facility can charge one vehicle at a time. We
refer to the vehicle that takes the \( i \)th charging facility as
vehicle \( i \). At stage \( t \), let \( I_t \) be the set of electric vehicles
that are at charging facilities, e.g., fast PHEV chargers.

3) For each vehicle \( i \in I_t \), let \( \alpha_i \) and \( \beta_i \) be its arrival
and departure time, respectively. Under the assumption
that the arrival and the departure occur at the beginning
of each stage, vehicle \( i \) can be charged from stage \( \alpha_i \)
through \( \beta_i - 1 \). We assume that \( 1 \leq \beta_i - \alpha_i \leq B \), i.e.,
each vehicle stays at a facility for at least one stage,
and for at most \( B \) stages.

4) At stage \( t \), for each vehicle \( i \in I_t \), let \( \gamma_{i,t} \) be the number
of time units (a nonnegative integer) needed to satisfy
vehicle \( i \)’s charging request. The charging rate of every
vehicle \( i \) is assumed to be constant from \( \alpha_i \) to \( \beta_i \) at the
facility. We let \( E \) be the maximum capacity of vehicles’
batteries, and therefore, \( \gamma_{i,t} \leq E \).

5) At stage \( t \), for each vehicle \( i \in I_t \), we use a two-
dimensional vector, \( x_{i,t} = (\lambda_{i,t}, \gamma_{i,t}) \) to denote the state of vehicle
\( i \), where \( \lambda_{i,t} = \beta_i - \gamma_{i,t} \) is a nonnegative integer
denoting the number of its remaining stages at facility
\( i \).

For notational convenience, if \( |I_t| = I < N \), we let
\( x_{i,t} = (0, 0) \), for \( i = I + 1, \ldots, N \). Here, and in
the sequel, we use \( |\cdot| \) to denote the cardinality of a particular
set.

6) For \( i \in I_t \), let \( a_{i,t} = 1 \) if vehicle \( i \) is charged at stage
\( t \); otherwise let \( a_{i,t} = 0 \).

7) A feasible action at stage \( t \), \( a_t = (a_{1,t}, \ldots, a_{N,t}) \), is an \( N \)-dimensional vector such that its \( i \)th component, \( a_{i,t} \), is either 1 (charging) or 0 (not charging), and satisfies
\( a_{i,t} \leq \gamma_{i,t} \). That is, vehicle \( i \) can be charged at stage
\( t \) if \( \gamma_{i,t} > 0 \) (it requires to be charged for at least one
stage).

8) Let \( A_t \) be the total number of vehicles charged at stage
\( t \), i.e., \( A_t = \sum_{i=1}^{N} a_{i,t} = \sum_{i \in I_t} a_{i,t} \).

9) For each vehicle \( i \in I_t \), it receives a utility \( a_{i,t} \cdot v \),
where \( v \) is a nonnegative constant.

10) At stage \( t \), let \( s_t \in S \) be the state of grid. We assume
that \( S \) is finite.

11) The cost associated with the battery charging of electric
vehicles is \( C(A_t, s_t) \).

12) The state of grid, \( s_t \), evolves as a Markov decision
process, where the transition probability depends on the
current grid state \( s_t \) and the number of vehicles charged
at stage \( t \), \( A_t \).

Note that charging a large number of vehicles may influ-
ence the Independent System Operator’s (ISO) dispatch
and reserve policy. To incorporate this type of impact,
we allow the evolution of the grid state to depend on
the aggregate action in general.

13) For a vehicle \( i \in I_t \), we have \( x_{i,t+1} = x_{i,t} - (1, a_{i,t}) \);
vehicle \( i \) is removed from the set \( I_t \) if and only if
\( \lambda_{i,t+1} \) becomes zero.

14) At stage \( t \), let \( d_t \in D \) be the state of demand, where
\( D \) is assumed to be finite. Let the number of arrival
vehicles at stage \( t+1 \) (i.e., those vehicles with an arrival
time \( t + 1 \) be a discrete random variable distributed according to \( \eta(dt) \), and that the initial states of the arriving vehicles (their departure times and charging requests) are independently and identically drawn according to a probability measure \( \xi(dt) \).

15) The states of demand, \( \{d_i\} \), evolves as a time-homogeneous Markov chain, and its state transition is assumed to be independent of the grid state and the action.

The state of grid, \( s_{t} \), and the state of demand, \( d_t \), are used to represent the exogenous factors, such as the current time, the aggregate demand of other electricity customers, and weather conditions, which have impacts on both the generators’ cost and the demand of PHEV owners. Our model allows the charging cost to depend on the exogenous state(s) \( s_{t} \) and therefore incorporates supply-side volatility due to the uncertainty in renewable generation. Note that the state transitions of \( \{s_t\} \) and \( \{d_t\} \) are assumed to be time-homogeneous. We next present a simple example to illustrate the state evolution of an electric vehicle.

Example 2.1: Consider a vehicle \( i \) that arrives at stage \( \alpha_i = 0 \), leaves at stage \( \beta_i = 8 \), and requires to be charged for 5 time units. Its initial state at stage 0 is therefore (8, 5). Suppose it is charged for one time unit during stage 0. Its state at stage 1 is (7, 4). Suppose that vehicle \( i \)'s state at stage 7 is (1, 2), and \( a_{i,7} = 1 \), i.e., it is charged at stage 7. At stage 8, vehicle \( i \) leaves the charging facility, and is removed from the set \( \mathcal{I}_8 \). However, its charging request is not satisfied before its departure: one additional time unit of charging (before its deadline, \( t = 8 \)) is needed to satisfy its request.

III. MARKOV DECISION PROCESS

In this section, we formulate the scheduling problem as an infinite-horizon Markov decision process (MDP) by providing the state space, action sets, transition probabilities, stage cost, and the average-cost objective function.

A. System State

At each stage \( t \), the system state, \( x_t \), consists of the states of all vehicles, \( \{x_{i,t}\}_{i=1}^N \), the grid state \( s_t \), and the state of demand \( d_t \). Let \( \mathcal{X} \) be the set of system states.

Note that the state space grows exponentially with the number of vehicles, \( N \). The cardinality of the system state space is \( (|B| + 1)(|E| + 1)^N \times |S| \times |D| \). Reasonable values of \( N, B, \) and \( E \) lead to very high dimensions, and make a direct solution of the MDP impossible.

B. Action

A feasible action at a system state \( x_t, a_t \), is an \( N \)-dimensional vector such that its ith component, \( a_{i,t} \), is either 1 (charging) or 0 (not charging) with \( a_{i,t} \leq \gamma_{i,t} \). We use

\[ U(x_t) \] to denote the set of feasible actions at the system state \( x_t \).

C. State Transition

The system state evolves as a Markov decision process, where the transition probability depends on the current system state, \( x_t \), and the current action \( a_t \). Since the state transition is independent of the stage \( t \), we use \( p_{x_t,y}(a) \) to denote the transition probability from state \( x \) to \( y \), under the action \( a \).

D. Cost

At each stage \( t \), the stage cost function, \( g(x_t, a_t) \), consists of three parts: the utility received by customers, \( v \sum_{i \in \mathcal{I}_t} a_{i,t} \), the cost \( C(A_t, s_t) \), and the penalty for unfilled charging requests of those vehicles that will leave at stage \( t + 1 \) (departure occurring at the beginning of each stage). Under the current system state \( x_t \) and the action \( a_t \), let \( J(x_t, a_t) \) be the set of unsatisfied customers (vehicles) that will leave charging facilities at stage \( t + 1 \), i.e.,

\[ J(x_t, a_t) = \{ j \mid \lambda_{j,t} = 1, \gamma_{j,t} - a_{j,t} > 0 \} \]

The stage cost function at stage \( t \) is

\[ g(x_t, a_t) = -v \sum_{i \in \mathcal{I}_t} a_{i,t} + C(A_t, s_t) + \sum_{j \in J(x_t, a_t)} q(\gamma_{j,t} - a_{j,t}) \]

where \( q(\cdot) \) reflects the utility loss (e.g., customers’ inconvenience and/or green house gas emission) due to the failure of satisfying PHEVs’ charging requests. Here, we have assumed that the penalty for an unsatisfied charging request (of vehicle \( j \)) depends only on the number of uncharged battery units \( (\gamma_{j,t} - a_{j,t}) \).

Note that the set of system states and the set of feasible actions are finite, and therefore the stage cost function is bounded. At a system state \( x_t \), \( g(x_t, a_t) \) can be regarded as the social welfare realized by the action \( a_t \) at stage \( t \).

E. Average Cost

Given an initial system state \( x_0 \), the time-averaged cost (from stage 0 through \( T - 1 \)) achieved by a policy \( \pi = \{v_{0}, v_{1}, \ldots\} \), where \( v_t(x_t) \in U(x_t) \), for all \( x_t \in \mathcal{X} \), is given by

\[ J_{\pi}^{T} (x_0) = \frac{1}{T} E_{\pi} \left\{ \sum_{t=0}^{T-1} g(x_t, v_t(x_t)) \right\} \]

where the expectation is over the distribution of future system states \( \{x_t\}_{t=1}^{T-1} \).

IV. LLLP PRINCIPLE

It is intuitive that the system state should incorporate the states of all vehicles. There are \( B(E + 1) + 1 \) possible states for a vehicle, i.e.,

\[ (\lambda_i, \gamma_i) \in \{(0,0), (1,0), (1,1), \ldots, (1,E), (2,0), (2,1), \ldots, (2,E), \ldots, (B,E)\}, \quad i = 1, \ldots, N. \]

Under the policy \( \pi \), the system state evolves as a Markov chain with transition probabilities \( p_{x_t, x_{t+1}}(v_t(x_t)) \).

2497
In this section, we first define a partial order over the set of \((B(E + 1) + 1)\) vehicle states listed on the right hand side of Eq. (3): a vehicle with less laxity and a longer remaining processing time has a higher-order state. For any given (possibly non-stationary) heuristic policy, we generate an “updated” policy (cf. the definition prior to Proposition 4.1) that gives priority to a vehicle with a higher-order state.

In Proposition 4.1, we show that an updated policy can only reduce the expected cost, compared with the original heuristic. At the end of this section, we introduce an “improved” policy generated from a given heuristic policy (based on the updated policy generated by the original heuristic; cf. the definition prior to Proposition 4.2), and discuss how to utilize the LLLP principle to improve a given heuristic policy.

At stage \(t\), based on vehicle \(i\)'s state \(x_{i,t} = (\lambda_{i,t}, \gamma_{i,t})\), we define
\[
\theta_{i,t} = \begin{cases} 
\lambda_{i,t} - \gamma_{i,t}, & \text{if } \gamma_{i,t} > 0, \\
B_i, & \text{if } \gamma_{i,t} = 0.
\end{cases}
\]

(4)

Note that for every vehicle \(i\) with \(\gamma_{i,t} > 0\), we have \(\theta_{i,t} \in \{1 - E, 2 - E, \ldots, B - 1\}\). For a vehicle \(i\) with \(\gamma_{i,t} > 0\), \(\theta_{i,t}\) reflects its laxity, i.e., the maximum number of stages that it can tolerate before the time it has to be put on continuous battery charging. A lower \(\theta_{i,t}\) implies a smaller flexibility to charge vehicle \(i\). For example, if \(\theta_{i,t} = 0\), then we have to keep charging vehicle \(i\) until its departure time to avoid the penalty for not satisfying its request; if \(\theta_{i,t} < 0\), it is impossible to satisfy vehicle \(i\)'s charging request before its departure.

Definition 4.1: At stage \(t\), for two vehicles \(i\) and \(j\) in the set \(I_t\), we say \(i \preceq j\) if \(\theta_{i,t} \geq \theta_{j,t}\) and \(\gamma_{i,t} \leq \gamma_{j,t}\). We say vehicle \(j\) is prior to vehicle \(i\) (\(i \prec j\)), if \(i \preceq j\) and \(x_{i,t} \neq x_{j,t}\).

It is not hard to see that the relation \(\preceq\) is reflexive, antisymmetric, and transitive, and therefore is a partial order. We also observe that if vehicle \(j\) has a less laxity and a later departure time than vehicle \(i\), i.e., if \(\theta_{i,t} \geq \theta_{j,t}\) and \(\lambda_{i,t} \leq \lambda_{j,t}\), then we must have \(i \preceq j\). At a system state \(x_t\), two vehicles \(i\) and \(j\) are incomparable, if \(\theta_{i,t} \geq \theta_{j,t}\) and \(\gamma_{i,t} > \gamma_{j,t}\), or \(\theta_{i,t} > \theta_{j,t}\) and \(\gamma_{i,t} \geq \gamma_{j,t}\). In that case, whether or not to give priority to a particular vehicle is a decision that depends on future grid states, as shown in the following example.

Example 4.1: At stage \(t\), suppose that we have two vehicles at charging facilities, i.e., \(|I_t| = 2\). The states of the two vehicles are \(x_{1,t} = (2, 1)\) and \(x_{2,t} = (4, 2)\). They are incomparable: vehicle 1 has a less laxity and a less remaining processing time. Suppose that no vehicles will arrive in the future. At stage \(t\), we have \(C(1, s_t) = 0\) and \(C(2, s_t) = 10\max\{v, q(3)\}\), where \(v\) is the consumer utility received from one unit battery charging. Obviously, it is optimal to charge only one vehicle at stage \(t\); however, which one to charge?

Consider two different scenarios after stage \(t\). First, suppose that the electricity is free at stages \(t + 2\) and \(t + 3\), but is expensive at stage \(t + 1\), say, \(C(A, s_{t+1}) = A\). It is easy to show that it is optimal to charge vehicle 1 at stage \(t\). On the other hand, if the electricity is free at stage \(t + 1\), but is expensive at stages \(t + 2\) and \(t + 3\), e.g., \(C(A, s_{t+2}) = C(A, s_{t+3}) = A\), then it is optimal to charge vehicle 2 at stage \(t\), and then charge both vehicles at stage \(t + 1\).

In the preceding example, at stage \(t\), even if vehicle 1 has less laxity than vehicle 2, it is not always optimal to charge vehicle 1, because the action of charging vehicle 1 (that has a less remaining processing time) at stage \(t\) restricts the set of feasible actions at stage \(t + 1\): we could at most charge one vehicle at stage \(t + 1\), if vehicle 1 were charged at stage \(t\). This also explains why among two vehicles with the same laxity, in Definition 4.1 priority is given to the vehicle with a longer remaining processing time (a later departure).

Before proving the main results of this paper, we introduce a convexity assumption for the penalty function.

Assumption 4.1: The penalty function \(q(\cdot)\) is nonnegative, and satisfies
\[
0 \leq q(n) - q(n - 1) \leq q(n + 1) - q(n), \quad n = 1, 2, \ldots
\]

For a given (possibly non-stationary) policy \(\pi = \{v_0, v_1, \ldots\}\), at some stage \(t\) and some system state \(x_t\), if there exists a pair of vehicles \(i, j \in I_t\) such that \(i \prec j\), and the policy \(\pi\) charges \(i\) and does not charge \(j\), we generate an updated policy \(\pi = \{v_0, \ldots, v_{t-1}, \nu_t, v_{t+1}, \ldots\}\) in the following way.

1) For \(k = t, t + 1, \ldots, v\), we first preset \(\nu_k = v_k\), and then follow the following procedure to update the sequence of mappings (from \(X\) to \([0, 1]^N\), \(\nu_k x_{k+1}\).

2) Policy \(\pi\) charges \(j\) instead of \(i\) at state \(x_t\). That is, \(\nu_t(x_t)\) is the same as \(v_t(x_t)\) except that its \(i\)th component is 0 and \(j\)th component is 1.

3) Let \(M = \max\{\lambda_{i,t}, \lambda_{j,t} - 1\}\) - 1. After the state \(x_t\), for any realization of system states that would occur with positive probability under the policy \(\pi\), \(\{x_k\}_{k=t+1}^{t+M}\), there exists a corresponding realization of system states after the state-action pair \((x_t, \nu_t(x_t))\), \(\{x_k\}_{k=t+1}^{t+M}\), such that the corresponding state \(x_{k}\) differs from \(x_{t}\) only on the states of vehicles \(i\) and \(j\): \(\gamma_{i,k} = \gamma_{i,k} + 1\) for \(k = t + 1, \ldots, \beta_i - 1\), and \(\gamma_{j,k} = \gamma_{j,k} - 1\) for \(k = t + 1, \ldots, \beta_j - 1\).

4) For every realization of system states that occurs with positive probability under the policy \(\pi\), \(\{x_k\}_{k=t+1}^{t+M}\), let \(\mathcal{G}(\{x_k\}_{k=t+1}^{t+M}) = \{t + 1, \ldots, \beta_i - 1\}\) be the set of stages that policy \(\pi\) charges vehicle \(i\), and does not charge vehicle \(i\), before vehicle \(i\)'s departure. If the set \(\mathcal{G}(\{x_k\}_{k=t+1}^{t+M})\) is empty, let \(\nu_k(x_k) = \nu_k(x_k)\), for \(k = t + 1, \ldots, t + M\).

If the set \(\mathcal{G}(\{x_k\}_{k=t+1}^{t+M})\) is not empty, let \(m\) be the minimal element in the set (note that \(m \leq \min\{\beta_i - 1, \beta_j - 1\}\)). At stages \(k = t + 1, \ldots, m - 1\), \(\nu_k(x_k) = \nu_k(x_k)\). At stage \(m\), policy \(\pi\) charges vehicle \(i\) instead.

6In Lemma 4.1, we will prove that for \(k = t + 1, \ldots, t + M\), whenever the policy \(\pi\) charges vehicle \(j\) at state \(x_k\), it is feasible to charge vehicle \(j\) at the corresponding state \(x_k\), i.e., \(\gamma_{j,k} \geq 1\).

Since \(i \prec j\) at state \(x_t\), and the policy \(\pi\) gives priority to vehicle \(i\) at state \(x_t\), we must have \(\gamma_{i,t+1} > \gamma_{i,t+1}\). For \(k = t + 1, \ldots, m - 1\), whenever the policy \(\pi\) charges vehicle \(j\) at state \(x_k\), it also charges vehicle \(i\). It follows that for \(k = t + 1, \ldots, m - 1\), if \(\pi\) charges vehicle \(j\) at state \(x_k\), then \(\gamma_{j,k} > \gamma_{i,k} \geq 1\), which implies that \(\gamma_{j,k} = \gamma_{j,k} - 1 \geq 1\).
of \( j \), i.e., \( \nu_m(\hat{x}_m) \) is the same as \( \nu_m(x_m) \) except that its \( i \)th component is 1 and its \( j \)th component is 0.

We will refer to the policy \( \pi \) defined above as the updated policy generated by the policy \( \pi \) at stage \( t \) on the state \( x_t \), with respect to vehicles \( i \) and \( j \) (or, more accurately speaking, with respect to the pair of states \( x_{i,t} \) and \( x_{j,t} \)). The following lemma justifies Step 5 in the definition of an updated policy.

**Lemma 4.1:** Suppose that at a system state \( x_t \), vehicle \( j \) is prior to vehicle \( i \), i.e., \( i \prec j \), and a policy \( \pi \) charges vehicle \( i \) and does not charge vehicle \( j \). For a given realization of system states resulting from the policy \( \pi \), \( \{x_k\}_{k=1}^{t+M} \), if the set \( G(\{x_k\}_{k=1}^{t+M}) \) is empty, then for \( k = t + 1, \ldots, t + M \), whenever the policy \( \pi \) charges vehicle \( j \) at state \( x_t \), we must have \( \hat{x}_{j,k} \geq 1 \), i.e., it is feasible to charge vehicle \( j \) at the corresponding state \( \hat{x}_k \).

We omit the proof of Lemma 4.1 for brevity. The intuition behind the definition of an updated policy is simple: an updated policy \( \hat{\pi} \) gives priority to vehicle \( j \) at stage \( t \), and will charge vehicle \( i \) at the first stage when the original policy \( \pi \) charges vehicle \( j \) and does not charge vehicle \( i \).

The following example illustrates this basic idea.

**Example 4.2:** At stage 0, suppose that we have two vehicles at charging facilities, i.e., \( |Z_0| = 2 \). The states of the two vehicles are \( x_{1,0} = (2,1) \) and \( x_{2,0} = (3,2) \). They are comparable: vehicle 2 is prior to vehicle 1. Now consider two heuristic policies:

1) Policy \( \pi \) charges only vehicle 1 at stage 0, charges only vehicle 2 at stage 1, and charges no vehicles at stage 2, regardless of future arrivals and future states of grid. For any realization of the system states, \( (x_1, x_2) \), the set \( G(x_1, x_2) \) is not empty and \( m = 1 \). An updated policy \( \hat{\pi} \) charges only vehicle 2 at stage 0, and at stage 1, it charges vehicle 1 instead of 2. At stage 2, it charges no vehicles.

2) Another policy \( \pi' \) charges only vehicle 1 at stage 0, and does not charge both vehicles 1 and 2 at stages 1 and 2, regardless of future arrivals and future states of grid. For any realization of the system states, \( (x_1, x_2) \), the set \( G(x_1, x_2) \) is empty. An updated policy \( \hat{\pi'} \) charges only vehicle 2 at stage 0, and does not charge vehicles 1 and 2 at stages 1 and 2.

In the preceding example, suppose that the state of grid allows the operator to charge at most one vehicle at stage 0. If priority is given to vehicle 1 that leaves earlier, the states of the two vehicles at stage 1 become \((1,0)\) and \((2,2)\); if the operator charges vehicle 2 instead of vehicle 1 at stage 0, the states of the two vehicles at stage 1 are \((1,1)\) and \((2,1)\). We argue that the latter situation (with vehicles’ states \((1,1)\) and \((2,1)\) at stage 1) is preferable, because (i) in the latter situation, the operator has a larger set of feasible actions at stage 1 (vehicle 1 is not available for charging in the former situation), and (ii) since the penalty function is convex, it is better to split remaining processing time into different vehicles. In general, since future states of grid are stochastic and the non-completion penalty is convex, it is preferable to have a larger number of vehicles with shorter remaining processing times. An updated policy gives priority to a vehicle with less laxity and longer processing time so as to keep a larger number of smaller tasks. The following proposition shows that an updated policy can only reduce the expected cost, compared to the original heuristic.

**Proposition 4.1:** Suppose that Assumption 4.1 holds. Let \( \pi = \{\nu_0, \nu_1, \ldots\} \) be a given policy, and let \( \hat{\pi} = \{\hat{\nu}_0, \hat{\nu}_1, \ldots\} \) be an updated policy generated by the policy \( \pi \) at a system state \( x_0 \). We have

\[
J_\pi^T(\hat{x}_0) \leq J_\pi^T(x_0), \quad \forall T \geq M + 1,
\]

where the time-averaged expected cost is defined in (2), and \( M = \max\{\lambda_j, 0, \lambda_i, 0\} - 1 \).

**Proof:** Suppose that at some system state \( x_0 \), vehicle \( j \) is prior to vehicle \( i \), and the policy \( \pi \) charges vehicle \( i \) and does not charge vehicle \( j \). Let \( \pi = \{\nu_0, \nu_1, \ldots\} \) be the updated policy (generated by the policy \( \pi \)) that charges vehicle \( j \) instead of \( i \) at state \( x_0 \). Note that the two policies, \( \pi \) and \( \pi' \), always charge an equal number of vehicles, and are identical after stage 2. Therefore, to prove the desired result, we only need to show that

\[
J_\pi^{M+1}(x_0) \leq J_\pi^M(\hat{x}_0).
\]

After the state \( x_0 \), for every realization of system states that would occur with positive probability under the policy \( \pi \), \( \{x_k\}_{k=1}^M \), there exists a corresponding realization of system states, \( \{\hat{x}_k\}_{k=1}^M \), which occurs with equal (positive) probability under the policy \( \pi \). Actually, if the set \( G(\{x_k\}_{k=1}^M) \) is empty, then \( \{x_k\}_{k=1}^M = \{\hat{x}_k\}_{k=1}^M \); otherwise, we have

\[
\{x_k\}_{k=1}^M = \{\hat{x}_0, \ldots, \hat{x}_m, \hat{x}_{m+1}, \ldots, \hat{x}_M\},
\]

where \( m \) be the minimal element in the set \( G(\{x_k\}_{k=1}^M) \). Therefore, to validate Eq. (5), it suffices to show that for every realization of system states under the policy \( \pi \), \( \{x_k\}_{k=1}^M \), and the corresponding realization of system states under the policy \( \hat{\pi} \), \( \{\hat{x}_k\}_{k=1}^M \),

\[
g(x_0, \nu_0(x_0)) + \sum_{k=1}^M g(x_k, \nu_k(x_k)) \leq g(x_0, \nu_0(x_0)) + \sum_{k=1}^M g(x_k, \nu_k(\hat{x}_k)).
\]

We now verify Eq. (6) by discussing four different cases:

1. For the realization of future system states, \( \{x_k\}_{k=1}^M \), there exists some stage \( k \leq \beta_i - 1 \) at which the policy \( \pi \) charges vehicle \( j \) and does not charge vehicle \( i \), i.e., the set \( G(\{x_k\}_{k=1}^M) \) is not empty; let \( m \) be the minimal element of the set. Note that \( m \leq \min\{\beta_i - 1, \beta_j - 1\} \). For the realization of system states \( \{x_k\}_{k=1}^M \), and the corresponding realization of system states under the policy \( \pi \), \( \{\hat{x}_k\}_{k=1}^M \), the two policies, \( \pi \) and \( \pi' \), always charge an equal number of vehicles, and result in the same cost from stage 0 through stage \( m \). At stage \( m + 1 \), the two policies, \( \pi \) and \( \pi' \), result in the same system state (i.e., \( \hat{x}_{m+1} = x_{m+1} \)), and policy \( \pi' \) will always agree with policy \( \pi \).
with the policy \( \pi (\nu_k = \pi_k \text{ for } k \geq m+1) \). For the pair of system state realizations, \( \{x_k\}_{k=1}^M \) and \( \{x_k\}_{k=1}^M \), we conclude that both policies result in the same cost, i.e., the equality holds in (6).

For the realization of system states, \( \{x_k\}_{k=1}^M \), let \( \rho_i \) denote additional time needed to satisfy vehicle \( i \)'s request when it leaves at stage \( \beta_i = \lambda_{i,0} \) under the policy \( \pi \). That is, \( \rho_i = \gamma_i, \beta_i - 1 - \alpha_i, \beta_i - 1 \), where \( \alpha_i, \beta_i - 1 \) is the \( i \)th component of \( \nu_{\beta_i - 1}(x_{\beta_i - 1}) \) (the action taken on vehicle \( i \) at state \( x_{\beta_i - 1} \) by the policy \( \pi \)). For the corresponding system state realization \( \{x_k\}_{k=1}^M \), we define \( \tilde{\pi}_i = \tau_i, \beta_i - 1 - \tilde{\pi}_i, \beta_i - 1 \), where \( \tilde{\pi}_i, \beta_i - 1 \) is the \( i \)th component of \( \tau_{\beta_i - 1}(x_{\beta_i - 1}) \), and \( \tilde{\pi}_i, \beta_i - 1 \) is the second component of vehicle \( i \)'s state at the system state \( x_{\beta_i - 1} \) (the action taken on vehicle \( i \) at state \( x_{\beta_i - 1} \) by the policy \( \tilde{\pi} \)). If the set \( G(\{x_k\}_{k=1}^M) \) is empty, then one of the following three cases must happen.

2. The strategy \( \pi \) keeps charging vehicle \( i \) until it leaves, i.e., the \( i \)th component of \( u_k(\nu_k) \) is 1, for \( k = 1, \ldots, \beta_i - 1 \). In that case, we must have \( \theta_{i,0} \leq 0 \), which implies \( \theta_{i,0} \leq \theta_{i,0} \leq 0 \). Under the policy \( \pi \), since vehicle \( i \) is charged from stage 0 through stage \( \beta_i \), and vehicle \( j \) is not charged at stage 0, we conclude that \( 0 \leq \rho_{i,0} < \rho_{j,0} \). Since at stages \( k = 1, \ldots, M \), \( u_k(\nu_k) = \tau_k(\nu_k) \), we have \( \tilde{\rho}_{j,0} = \rho_{j,0} - 1 \geq 0 \) and \( \tilde{\rho}_{i,0} = \rho_{i,0} + 1 \geq 1 \). According to Assumption 4.1 we have

\[
q(\tilde{\rho}_{j,0}) + q(\tilde{\rho}_{i,0}) \leq q(\rho_{j,0}) + q(\rho_{i,0}).
\]

Note that the two policies, \( \pi \) and \( \tilde{\pi} \), result in the same cost except possible different penalties for not satisfying vehicle \( i \)'s and \( j \)'s charging requests. Therefore, Eq. (7) implies the desired result in (6).

3. Vehicle \( i \) leaves no later than \( j \) (\( \beta_i \leq \beta_j \)), and whenever \( \pi \) does not charge vehicle \( i \), it does not charge vehicle \( j \) either. It can be shown that Eq. (7) holds in this case.

4. Vehicle \( j \) leaves no later than \( i \) (\( \beta_j \leq \beta_i \)), and before vehicle \( j \)’s departure, whenever \( \pi \) does not charge vehicle \( i \), it does not charge vehicle \( j \) either. Eq. (7) also holds for this case.

Although an updated policy cannot be worse than the original heuristic policy \( \pi \), it may not fully utilize the extra feasibility provided by the LLLP principle. As an example, consider the first heuristic policy \( \pi \) and its updated policy \( \tilde{\pi} \) in Example 4.2. At stage 1, under the updated policy \( \tilde{\pi} \), the states of vehicles 1 and 2 are \( (1,1) \) and \( (2,1) \), respectively. According to the the LLLP principle, the updated policy charges vehicle 2 instead of vehicle 1 at stage 0, and expands the feasible action set at stage 1 (under the original policy \( \pi \), vehicle 1 would not be available for charging at stage 1). At stage 1, the updated policy charges vehicle 1 instead of vehicle 2. If the electricity is free at stage 1, and if charging one more vehicle at stage 1 has no influence on the evolution of the grid state, then it is better to charge both vehicles 1 and 2 at stage 1, regardless of future PHEV arrivals and future grid states. This observation motivates us to design a policy that is allowed to charge more vehicles than the original heuristic. We next define a simple “improved” policy that charges more vehicles than the original heuristic, when necessary.

For a given policy \( \pi \), we define an improved policy of the original heuristic policy \( \pi \) at stage \( x_t \) with respect to vehicles \( i \) and \( j \), \( \tilde{\pi} = (\nu_0, \nu_1, \nu_i, \nu_{i+1}, \ldots) \), as follows. The improved policy agrees with the corresponding updated policy \( \tilde{\pi} \) except that for a realization of system states (under the original policy \( \pi \)) \( \{x_k\}_{k=1}^{M+1} \) such that the set \( G(\{x_k\}_{k=1}^{M+1}) \) is not empty, the policy \( \tilde{\pi} \) charges both vehicles \( i \) and \( j \) at the corresponding system state \( x_m \) (\( m \) is the minimal element of the set \( G(\{x_k\}_{k=M+1}) \)), if one of the following conditions holds:

1) \( \theta_{j,m} \leq 0 \) and \( C(A_m + 1, s_m) - C(A_m, s_m) < q(1) \), where \( A_m \) is the number of charged vehicle under the policy \( \pi \) at state \( x_m \). That is, vehicle \( j \) has a nonpositive laxity and the penalty for not satisfying its request is larger than the incremental charging cost at stage \( m \).

2) \( C(A_m + 1, s_m) = C(A_m, s_m) \), i.e., it is free to charge one more vehicle at stage \( m \).

Although the preceding two conditions seem to be somewhat restrictive, the definition of an improved policy sheds some light on how to improve a heuristic policy \( \pi \): the new policy agrees with the updated policy \( \tilde{\pi} \) except that it charges both vehicles \( i \) and \( j \) at stage \( m \) (the minimal element of the set \( G(\{x_k\}_{k=M+1}) \)), if it is not empty, if the additional cost of charging one more vehicle is reasonably small, compared with the (forecasted) charging cost in the future and the potential risk of not satisfying vehicle \( j \)'s request before its departure.

Proposition 4.2: Suppose that Assumption 4.1 holds, and that the evolution of the grid state is independent of the operator’s action. Let \( \pi = (\nu_0, \nu_1, \nu_{i+1}, \ldots) \) be a given policy, and let \( \tilde{\pi} = (\tilde{\nu}_i, \tilde{\nu}_{i+1}, \ldots ) \) be an improved policy generated by \( \pi \) at a system state \( x_t \). We have

\[
J^T_T(x_t) \leq J^T_T(x_t), \quad \forall T \geq M + 1,
\]

where \( M = \max(\lambda_{i,t}, \lambda_{j,t}) - 1 \).

Since the evolution of the grid state is exogenous, it can be shown that \( J^T_T(x_t) \leq J^T_T(x_t) \), for any \( T \geq M + 1 \). Proposition 4.2 then follows from Proposition 4.1.

V. APPLICATION

In this section, we compare the performance of three stationary heuristic policies, the EDF (Earliest Deadline First) policy, and two LLF-based (Least Laxity First) heuristic policies. We assume that the states of grid, \( \{s_0, s_1, \ldots \} \), are independent and identically random variables uniformly distributed over the set \( \mathcal{S} = \{40, 41, \ldots, 160\} \). The state of grid reflects the capacity available for charging PHEV batteries. The penalty function is assumed to be linear, i.e., \( q(n) = n \). For simplicity, we let \( v = 0 \), and assume that the electricity is free (an ideal model for the case where the charging cost is much smaller than the penalty for not satisfying customers’ charging requests). That is, for a given \( s \in \mathcal{S} \), the cost function is given by

\[
C(A, s) = 0, \quad \text{if } A \leq s; \quad C(A, s) = Nq(N), \quad \text{if } A > s,
\]
which implies that the operator should never charge more than \( s_t \) vehicles at stage \( t \). Note that in our simulation, the stage cost (defined in (1)) depends only on the penalty for unsatisfied charging requests.

We assume that the number of arriving vehicles is a time-invariant constant, and that the initial states of arriving vehicles are independent and identically distributed random variables. We assume that the number of stages for which a newly arrived vehicle \( i \) will stay at a charging facility, \( \beta_i - \alpha_i \), is uniformly distributed over the set \([1, \ldots, 10]\) (i.e., \( B = 10 \)), and that the time needed to fulfill its request, \( \gamma_{i,a} \), is uniformly distributed over the set \( \{1, \ldots, 10\} \).

For a system state \( x_t \), let \( V(x_t) \) be the number of vehicles in the set \( I_t \) that require to be charged for at least one stage. At a system state \( x_t \), the stationary EDF policy, \( \mu \), charges the first \( \min\{s_t, V(x_t)\} \) vehicles that have the earliest departure times. For two vehicles that have the same departure time, the policy \( \pi \) charges the one with less laxity.

We compare the performance of EDF policy with two LLF-based (Least Laxity First) stationary policies. At a system state \( x_t \), both LLF-based policies charge the first \( \min\{s_t, V(x_t)\} \) vehicles that have the least laxity. For two vehicles with the same laxity, the LLSP (Least Laxity and Shorter remaining Processing time) policy, \( \mu^L \), gives priority to the vehicle that has a shorter remaining processing time (an earlier departure time), while the LLLP (Least Laxity and Longer remaining Processing time) policy, \( \mu^{LL} \), gives priority to the vehicle that has a longer remaining processing time.

We let \( N = 400 \), which is large enough to accept all arriving vehicles in our simulation. The average cost resulting from the three heuristic policies, EDF, LLSP and LLLP, are compared in Fig. 1. The numerical results show that the LLLP policy achieves the lowest time-averaged cost, and the performance of LLSP is better than EDF, and that with mild system overload, say, when the arrival rate is less than 30, the LLLP policy reduces the time-averaged cost by 15\% – 35\%, compared to the LLSP policy.

VI. CONCLUSION

We formulate the scheduling problem of PHEV charging as an infinite-horizon Markov decision process, with an objective to maximize the time-averaged expected social welfare. We show the less laxity and longer remaining processing time (LLLP) principle: priority should be given to vehicles that have less laxity and longer remaining processing times, if the non-completion penalty function is convex, and if the operator does not discount future cost. Based on the LLLP principle, we introduce two forms of polices that are generated from a given heuristic policy, and show (under mild assumptions) that the generated polices can only reduce the expected cost, compared to the original heuristic. Numerical results demonstrate that heuristic policies that violate the LLLP principle (e.g., the EDF policy and the LLSP policy) may result in significant losses of social welfare.

REFERENCES