Large Scale Charging of Electric Vehicles: A Multi-Armed Bandit Approach

Zhe Yu†, Yunjian Xu‡, and Lang Tong†

***Abstract*—The problem of centralized scheduling of large scale charging of electric vehicles (EVs) by a service provider is considered. A Markov decision process model is introduced in which EVs arrive randomly to the charging facility with random demand and completion deadlines. The service provider faces random charging costs, convex non-completion penalties, and a peak power constraint that limits the maximum number of simultaneous activation of EV chargers.**

Formulated as a restless multi-armed bandit problem, the EV charging problem is shown to be indexable, thus low complexity index policies exist. A closed-form expression of the Whittle's index is obtained for the case when the charging costs are constant. The Whittle's index policy, however, is not optimal in general. An enhancement of the Whittle's index policy based on spatial interchange according to the less laxity and longer processing time (LLLP) principle is presented. The proposed policy outperforms existing charging algorithms, especially when the charging costs are dynamic.

***Index Terms*—Multi-armed bandit problem; deadline scheduling; charging of electric vehicles; Whittle's index; Markov decision processes.**

I. Introduction

ELECTRIC vehicles (EVs) and EV charging services have grown substantially. Since 2011, the EV sale has grown 20 times, and EV charging stations have increased 7.5 folds [1], [2]. Although most EVs in use currently are charged at private homes, there is a need to develop large charging facilities with fast charging capabilities in public spaces such as parking garages, parking lots at commercial locations, and highway rest stops. Such charging facilities alleviate range anxiety of EV consumers; they are essential to the growth of EV market share [3].

EV charging at facilities with capacity of hundreds of EVs faces a different set of technical challenges from those associated with individual home charging. First, consumers expect charging to be completed within a relatively short period of time. Thus, fast charging devices operated at high peak power becomes essential. This type of charging, if unmanaged, may have detrimental effects on system reliability. It may be necessary to limit the number of simultaneously activated chargers.

Second, there is significant uncertainty in charging demand. EVs arrive at a charging facility randomly, each with stochastic demand and random deadlines, which makes it difficult for the scheduler to meet consumer demands.

Third, the cost (or the profit) of the service provider may be stochastic. For instance, the service provider may participate in the wholesale electricity market and is subject to real-time price fluctuations. In addition, the service provider may integrate local renewables such as solar generations that are intermittent.

Finally, the energy management system that schedules EV charging needs to operate in real time, thus must be scalable with respect to the size of the charging facility, which rules out the use of brute-force optimization techniques.

A. Summary of results

This paper presents an online scheduling algorithm that is computationally scalable and capable of dealing with demand and cost uncertainty. We introduce a constrained Markov decision process (MDP) model with the objective of maximizing expected (discounted) profit subject to a constraint on the maximum number of simultaneously activated chargers. The model captures randomness in arrivals, demand, deadlines, and charging costs. The optimal charging problem is then reformulated as a restless Multi-Armed Bandit problem (MAB) with simultaneous plays [4]. We establish the indexability for the restless MAB problem. For the constant cost case, we obtain the Whittle's index in closed form, which exposes certain weaknesses of Whittle's index policy. Because Whittle's index policy is not optimal in general, we present an improvement of the Whittle's index policy using a spatial interchange procedure based on the less laxity and longer processing (LLLP) time principle [5]. Numerical simulations demonstrate improved performance, especially when the cost of charging is dynamic and traffic is relatively heavy.

B. Related work

The centralized EV charging problem considered in this paper falls in the category of *stochastic multi-processor deadline scheduling problem*. In that context, EVs are jobs and chargers are processors. The work most relevant to the current paper is [6] by Raghunathan, Bokar, and Kumar on a deadline scheduling problem in wireless communications. The authors of [6] are perhaps the first to formulate the stochastic deadline scheduling problem as a restless MAB problem and established indexability. Also related is [7] where the problem of scheduling packets with deadlines in ad hoc networks is considered.

Z. Yu† and L. Tong† are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA. Y. Xu‡ is with the School of Engineering Systems and Design Pillar, Singapore University of Technology and Design, Singapore, 487372. Email: {zy73,lt35}@cornell.edu, yunjian_xu@sutd.edu.sg.
This work is supported in part by the National Science Foundation under Grant CNS-1248079

There are several nontrivial differences between the models in [6], [7] and that in the current paper. For instance, the arrival models used in [6] are either simultaneous or periodic. The cost models in [6], [7] are also significantly different from ours. The results in [6], [7] do not apply directly here.

The dynamic programming approach to EV charging was considered in [5] where the Less Laxity and Longer Processing time (LLLP) principle was first established. LLLP is an enhancement of any policy via a spatial interchange argument, and it is used in this paper on the Whittle's index policy. Earlier, Panwar and Towsley considered the case when there is a single processor [8], [9]. It was shown that the earliest deadline first (EDF) scheduling minimizes the unfinished work. See also the more recent work of Lehoczky and Shreve [10]. The performance of stochastic deadline scheduling problem involving multiple processors is largely unknown.

The deadline scheduling problem was originally considered by Liu and Layland [11] and studied extensively under models with deterministic arrivals, job sizes, and deadlines. See [12], [13] for some of the classical results. The single processor case is relatively well understood. For the multiprocessor problem, the results are limited. In fact, optimal online scheduling does not exist in general [14].

Applications of deterministic deadline scheduling to EV charging are considered in [15], [16], [17], [18]. Because arrivals and demands are arbitrary, performance measures used are often based on the worst case scenarios, which may not be appropriate for evaluating the performance of EV charging services.

Different from the centralized scheduling of EV charging framework considered here, there is a significant body of literature on home EV charging problems, often in decentralized optimization or game theoretic settings. For example, the work in [19] and [20] aims to minimize the load variance. Under a game theoretic model, the authors showed that the Nash Equilibrium is the optimal "valley filling" policy. A decentralized control algorithm for the same purpose is proposed in [21]. The vehicle to grid (V2G) ancillary services are considered in [22], [23], [24]. In [22], an optimal pricing strategy is proposed to encourage EVs to participate in frequency regulation and a centralized control is used in [24].

II. Problem Formulation

We now formulate the EV charging problem as one of stochastic deadline scheduling subject to a constraint on the number of servers (chargers). In Section II-A, a constrained Markov Decision Processes (MDPs) problem is introduced. In Section II-B, we provide an upper bound on the total discounted reward, which is useful for benchmark comparisons.

A. An MDP formulation of Stochastic Deadline Scheduling

Fig. 1 shows a schematic of an energy management system at an EV charging facility. We assume that the facility has N parking spots, each with a charger that can be activated or deactivated by the scheduler. The charger can only be connected to one vehicle. It is *open* if it is not attached to a vehicle and *closed or occupied* otherwise.

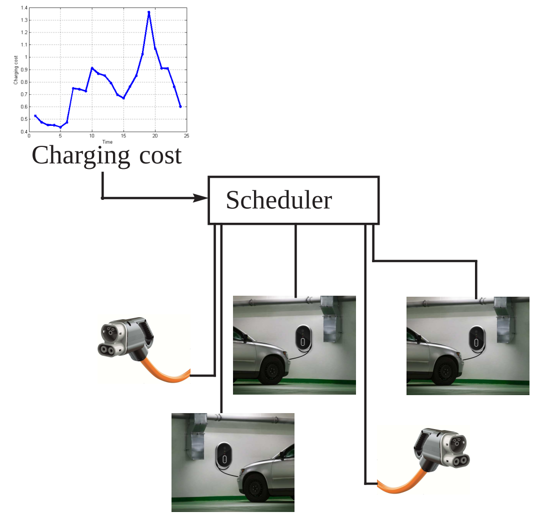

Fig. 1: Architecture of a charging station

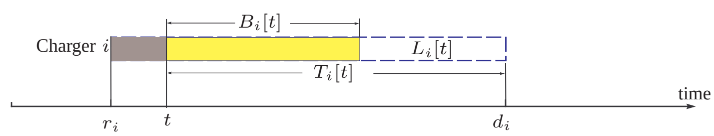

Fig. 2: An illustration for the charger's state. r_i is the arrival time of an EV at charger i, d_i the deadline for completion, $B_i[t]$ the amount of charging to be completed by d_i, $T_i[t]$ the lead time to deadline.

EVs arrive at chargers independently. At the arrival of charger i, if the charger is open, the EV is attached to it, and the scheduler records the arrival time r_i, as shown in Fig. 2. The EV owner communicates the charging demand B_i, measured in charging time, and deadline for completion d_i to the scheduler, also shown in Fig. 2.

We summarize the assumptions in the paper; they are approximations of practical operating conditions and are made for tractable analytical developments.

- A1. Each charger can be connected to only one EV, and it is removed from the EV at the deadline d_i. An EV is charged at a fixed rate normalized to 1.
- A2. The EV arrivals to the N chargers are independent and identically distributed (i.i.d.).
- A3. The price of charging is proportion to the charging demand, normalized to 1 dollar/hour.
- A4. The marginal charging cost $c[t]$ is an exogenous finite state Markov chain whose evolution is independent of the state evolution and actions of charging.
- A5. The charging of EVs is preemptive without cost.
- A6. The penalty for incomplete charging is a convex function of the incomplete amount at the deadline.

We now present elements of the discrete-time MDP in which time, indexed by $t = 0, 1, 2, \cdots$, is slotted. At the beginning of the slot, the system state is revealed to the scheduler and a decision on which chargers to activate or deactivate in the current slot is made and executed.

1) State space: The state of the charging system

$$S[t] = (c[t], S_1[t], \cdots, S_N[t]) \in \mathcal{S}_c \times \mathcal{S}_1 \times \cdots \times \mathcal{S}_N$$

is defined by the charging cost $c[t]$ and states of individual chargers $S_i[t]$ where $\mathcal{S}_c$ is the state space of the cost and $\mathcal{S}_i$ the state space of individual chargers. Specifically, the state of charger i is defined by $S_i[t] \triangleq (T_i[t], B_i[t])$ where, as illustrated in Fig. 2, $T_i[t] \triangleq d_i - t$ is the lead time and $B_i[t]$ the remaining charging demand measured in charging time. If there is no EV attached to charger i, then $S_i[t] = (0, 0)$. The system charging cost $c[t]$ is the cost of electricity from the wholesale market, offset by possibly locally generated renewables.

2) Action: The action of the scheduler is defined by $a[t] = (a_1[t], \cdots, a_N[t]) \in \{0, 1\}^N$ where $a_i[t] = 1$ means that the charger is activated (active) whereas $a_i[t] = 0$ means that the charger is deactivated (passive).

3) State evolution: We assume that the charging cost $c[t] \sim (\mathcal{S}_c, P)$ is an exogenous finite state Markov chain, independent of the actions of the scheduler and individual state evolutions, with transition probability matrix $P = [P_{i,j}]$.

Given the scheduled action $a[t] = (a_i[t])$, the evolution of states at individual chargers are assumed statistically independent. When the charger is active and the vehicle has positive remaining demand, both the charging demand and the lead time are reduced by 1. If the charging demand of an EV is fulfilled ($B_i[t] = 0$), then only the lead time is decreased by one. EVs leave at their deadlines and new EVs arrive following a geometric distribution and the state probability mass function (PMF) $Q(\cdot, \cdot)$.

Specifically, the state of charger i with state $S_i[t]$ under action $a_i[t] = 1$ is transitioned to

$$\Big(S_i[t+1] \mid a_i[t] = 1\Big) = \begin{cases} (T_i[t]-1, B_i[t]-1) & \text{if } B_i[t] > 0, T_i[t] > 1, \\ (T_i[t]-1, B_i[t]) & \text{if } B_i[t] = 0, T_i[t] > 1, \\ (0,0) & \text{w.p. } (1-\rho), \quad \text{if } T_i[t] \leq 1, \\ (1,1) & \text{w.p. } \rho Q(1,1), \quad \text{if } T_i[t] \leq 1, \\ \cdots & \\ (T_{\max}, B_{\max}) & \text{w.p. } \rho Q(T_{\max}, B_{\max}), \ \text{if } T_i[t] \leq 1, \end{cases} \tag{1}$$

where ρ is the probability that an EV arrives at an empty charger and $Q(\cdot, \cdot)$ the PMF of the state of a newly arrived EV.

Similarly, the state of charger i under the passive action $a_i[t] = 0$ has the state transition given by

$$\Big(S_i(t+1) \mid a_i[t] = 0\Big) = (T_i[t] - 1, B_i[t]) \tag{2}$$

4) Reward: At time t, the reward received from charger i with action a is given by

$$R_a(S_i[t], c[t]) = \begin{cases} (1-c[t])a, & \text{if } B_i[t] > 0, \ T_i[t] > 1, \\ (1-c[t])a & \\ \quad - F(B_i[t]-a), & \text{if } B_i[t] > 0, \ T_i[t] = 1, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where $F(B)$ is an increasing and convex penalty function with $F(0) = 0$. Note that the above reward function means that the EV owner with charging demand B is charged B dollars if the charging request is fulfilled at the deadline and $B - B' - F(B')$ if there is B' unfulfilled charging. Here $F(B')$ is the extra compensation for unfulfilled charging.

Given the initial system state $S[0] = s$ and a policy π that determines a sequence of actions $a[t], t = 0, 1, \cdots$, the expected discounted system reward is defined by

$$V_\pi(s) \triangleq \mathbb{E}_\pi \left(\sum_{t=0}^{\infty} \sum_{i=1}^{N} \beta^t R_{a_i[t]}(S_i[t], c[t]) \middle| S[0] = s \right) \tag{4}$$

where $\mathbb{E}_\pi$ is the conditional expectation for given scheduling policy π and $0 < \beta < 1$ the discount factor.

5) Constrained MDP and optimal policy: We impose a constraint on the number of simultaneously activated chargers, $\sum_i^N a_i[t] \leq M$ for all t. In practice, such a constraint limits the peak power consumption by the service provider, which is necessary to satisfy the feeder and line limits.

The optimal EV charging policy can now be formulated as a constrained MDP. Specifically, let

$$V(s) = \sup_{\{\pi: \sum_i^N a_i^\pi[t] \leq M \quad \forall t\}} V_\pi(s) \tag{5}$$

where $a_i^\pi[t]$ is the action generated by π. A policy π^* is optimal if $V_{\pi^*}(s) = V(s)$.

Without loss of generality, we restrict ourselves to the set of stationary policies [25].

B. A performance upper bound

In (5), the power limit must be satisfied for all t. By relaxing this constraint and requiring that the average power usage does not exceed M, we obtain a performance upper bound for (5). In particular, a relaxed problem can be stated as

$$\begin{array}{ll} \sup_\pi & \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \sum_{i=1}^{N} \beta^t R_{a_i[t]}(S_i[t], c[t]) \mid S[0] \right\} \\ \text{subject to} & (1-\beta)\mathbb{E} \sum_{t=0}^{\infty} \sum_{i=1}^{N} \beta^t a_i[t] \leq M. \end{array} \tag{6}$$

Problem (6) is not a practical formulation of the large scale EV charging problem since the power usage could be far more than M at some time.

Since the charging cost is the same for all chargers, the relaxed problem (6) is equivalent to the following problem (on the scheduling of a single charger i).

$$\begin{array}{ll} \sup_\pi & N\mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \beta^t R_{a_i[t]}(S_i[t], c[t]) \mid S_i[0], c[t] \right\} \\ \text{subjec to} & (1-\beta)\mathbb{E} \sum_{t=0}^{\infty} \beta^t a_i[t] \leq M/N. \end{array} \tag{7}$$

Problem (7) is to maximize the discounted reward from a single charger i with no more than M/N active action (per time period) on average. The optimal solution and the optimal objective of (7) are the same as those of (6). The optimal objective of (7) can be used as a performance upper bound for the original scheduling problem in (5).

The constrained MDP problem in (7) has a much smaller dimension and can be easily solved by linear programming (cf. Chap. 3 of [25] for a survey).

III. Index policy and Whittle's index

The MDP formulation does not result in a scalable scheduling policy. Because charging at individual chargers is independent conditioned on charging cost, we seek to obtain an *index policy* that provides a scalable solution. By index policy we mean that the scheduling is based on the ranked order of indices associated with chargers. Specifically, the index of charger i is a mapping from its extended state $\tilde{S}_i[t] \triangleq (S_i[t], c[t])$ to a index value.

A. Deadline scheduling as a restless MAB problem

We now formulate Problem (5) as a restless Multi-Armed Bandit (MAB) problem. The restlessness is due to the fact that the lead time of each charger evolves even if the charger is not activated.

A complication of casting (5) as a restless MAB problem comes from the inequality constraint on the maximum number of simultaneous activations. This complication can be circumvented by introducing M dummy chargers to the scheduling problem. Specifically, each dummy charger accrues zero reward regardless of the actions applied to it, and the state of dummy chargers stays at $S_i = (0, 0)$. Without loss of generality, let $i \in \{1, \cdots, N\}$ be the regular chargers and $i \in \{N+1, \cdots, N+M\}$ be the dummy chargers.

1) Arms: We let each charger be an arm. Define the extended state of each charger as $\tilde{S}_i[t] \triangleq (S_i[t], c[t])$ and denote the extended state space as $\tilde{\mathcal{S}}_i \triangleq \mathcal{S}_i \times \mathcal{S}_c$. The actions and the reward functions remain unchanged.

For the regular chargers, since the cost dynamic is independent of the state and actions of charging, the state transition of arm i can be easily written according to charger transition (1), (2) and cost transition P. For the dummy chargers, only the charging cost evolves according to P.

2) MAB formulation: By including dummy chargers, the MDP in (5) can be viewed as a restless MAB problem that, at each time, exactly M out of $N+M$ chargers (arms) are active. The optimization problem with the equality constraint is state as following:

$$\begin{array}{ll} \sup_\pi & \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \sum_{i=1}^{N+M} \beta^t R_{a_i[t]}(\tilde{S}_i[t]) \mid \tilde{S}_i[0] \right\} \\ \text{s.t.} & \sum_{i=1}^{N+M} a_i[t] = M, \quad \forall t. \end{array} \tag{8}$$

It can be shown that the optimization with the inequality constraint defined Problem (5) is equivalent to the MAB involving dummy chargers with the equality constraint in (8).

B. Whittle's index

We now examine the Whittle's index policy for the restless MAB problem defined in (8). To this end, we first introduce Whittle's index and establish in Sec III-C the indexability of the restless MAB problem in Theorem 1.

Consider the following single arm reward maximizing problem without constraint: given the initial state $\tilde{S}_i[0]$,

$$V_i(\tilde{s}) \triangleq \sup_\pi \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \beta^t R_{a_i[t]}(\tilde{S}_i[t]) \mid \tilde{S}_i[0] = \tilde{s} \right\}, \tag{9}$$

where V_i is the value function. Let $\mathcal{L}_a$ be the Markov transition operator on the extended state $\tilde{S}_i$ and an arbitrary function $f(\tilde{S}_i)$ defined as

$$(\mathcal{L}_a f)(\tilde{s}) \triangleq \mathbb{E}\{f(\tilde{S}_i[t+1]) \mid \tilde{S}_i[t] = \tilde{s}, a_i[t] = a\}.$$

The maximum discounted reward of Problem (9) is determined by the Bellman equation

$$V_i(\tilde{s}) = \max\{R_0(\tilde{s}) + \beta(\mathcal{L}_0 V_i)(\tilde{s}), R_1(\tilde{s}) + \beta(\mathcal{L}_1 V_i)(\tilde{s})\}.$$

The Whittle's index is defined by introducing a subsidy ν paid to the scheduler to take the passive action [4]. The Bellman equation for the ν-subsidy problem is given by

$$V_i^\nu(\tilde{s}) = \max\{R_0(\tilde{s}) + \nu + \beta(\mathcal{L}_0 V_i^\nu)(\tilde{s}), R_1(\tilde{s}) + \beta(\mathcal{L}_1 V_i^\nu)(\tilde{s})\},$$

where V_i^ν is the value function for the ν-subsidy problem.

Intuitively, the larger the subsidy ν is, the more likely the passive action would be optimal. Let $\tilde{\mathcal{S}}_i(\nu)$ denote the set of arm states in which it is optimal to take the passive action on arm i in the ν-subsidy problem. The indexability of an MAB problem is defined as follows.

Definition 1 (Indexability): Charger (arm) i is indexable if the set $\tilde{\mathcal{S}}_i(\nu)$ increases monotonically from $\emptyset$ to $\tilde{\mathcal{S}}_i$ as ν increases from $-\infty$ to $+\infty$. The MAB problem is indexable if all the chargers (arms) are indexable.

Given the definition of indexability, the Whittle's index is defined as follows.

Definition 2 (Whittle's Index): If charger (arm) i is indexable, its Whittle's index $\nu_i(\tilde{s})$ of the extended state $\tilde{s}$ is the infimum subsidy ν such that the passive action is optimal at state $\tilde{s}$, *i.e.*,

$$\nu_i(\tilde{s}) \triangleq \inf_\nu \{\nu : R_0(\tilde{s}) + \nu + \beta(\mathcal{L}_0 V_i^\nu)(\tilde{s}) \geq R_1(\tilde{s}) + \beta(\mathcal{L}_1 V_i^\nu)(\tilde{s})\}.$$

C. Indexability and index closed-form

In this subsection, we apply the Whittle's index to chargers (arms) in problem (8). In the following theorem, we establish the indexability of each charger; for the case with constant charging cost, we also derive the closed-form expression of Whittle's index.

Theorem 1 (Indexability and index closed-form):
1) Each charger as an arm is indexable.
2) If $c[t] = c_0$ for all t, Whittle's index of a regular charger $i \in \{1, \cdots, N\}$ is given by

$$\nu_i(T,B,c_0)=\begin{cases}0 & \text{if } B=0,\\ 1-c_0 & \text{if } 1\le B\le T-1,\\ 1-c_0+\beta^{T-1}[F(B-T+1)-F(B-T)] & \\ & \text{if } T\le B.\end{cases} \quad (10)$$

Whittle's index of a dummy charger is zero.

$$\nu_i(0,0,c_0)=0, i\in\{N+1,\cdots,N+M\}.$$

In (10), when it is feasible to fulfill the EV's charging request, its Whittle's index is simply the charging profit $1-c_0$. When the penalty is inevitable, the index takes into account both the charging profit and the non-completion penalty. We note that the Whittle's index gives high priority to urgent EVs with non-positive laxity. Here, the laxity of charger i is defined as $L_i[t]\triangleq T_i[t]-B_i[t]$ (cf. Fig. 2). We note, however, that the Whittle's index does not distinguish EVs with positive laxity. In the next section we will introduce an enhanced heuristic policy based on the Whittle's index.

The proof of Theorem 1 is omitted due to the space limit. A proof sketch is provided in the following.

1) Dummy Chargers: For $i\in\{N+1,\cdots,N+M\}$, there is no EV arrival, and only the charging cost evolves. The Bellman equation of the ν-subsidy problem is given by

$$\begin{aligned}V_i^{\nu}(0,0,c_j) \quad = \quad & \max\{\beta\textstyle\sum_k P(j,k)V_i^{\nu}(0,0,c_k),\\ & \nu+\beta\textstyle\sum_k P(j,k)V_i^{\nu}(0,0,c_k)\}.\end{aligned}$$

When $\nu<0$, it is optimal to activate the dummy charger. Otherwise, passive action is optimal. So a dummy charger is indexable and its Whittle's index is $\nu_i(0,0,c_j)=0$, according to Definition 2.

2) Regular Chargers: For $i\in\{1,\cdots,N\}$, the indexability can be proved by induction. We first find out Whittle's index for a charger with $T_i\le 1$. Assuming Whittle's index exists for T_i, we show that for T_i+1, the difference between active and passive actions is monotonically increasing in ν and has a zero point. Thus Whittle's index exists for T_i+1.

IV. Whittle's Index Policy with LLLP Interchange

For the objective of time average ($\beta=1$) profit maximization, Whittle's index policy is shown to be asymptotically optimal, as the number of arms increases to infinity [26], under some conditions. For the discounted profit maximization setting considered in this paper, the asymptotic optimality of Whittle's index policy is not clear. For small systems with finitely many arms, there are counter-examples where an optimal index policy does not exist (and therefore the Whittle's index policy cannot be optimal). In this section, we will apply the Less Laxity and Longer Processing time (LLLP) principle (proposed in [5]) to improve the Whittle's index policy.

A. Less Laxity and Longer Processing time principle

The LLLP principle is a priority rule for the scheduling of charging multiple EVs, which is defined as follows.

Definition 3 (LLLP Priority): Consider chargers (arms) i and j at time t. We say j dominates i ($j\succeq i$), if j has Less Laxity and Longer Processing time, *i.e.,* $B_j[t]\ge B_i[t]$ and $L_j[t]\le L_j[t]$, with at least one of the inequalities being strict.

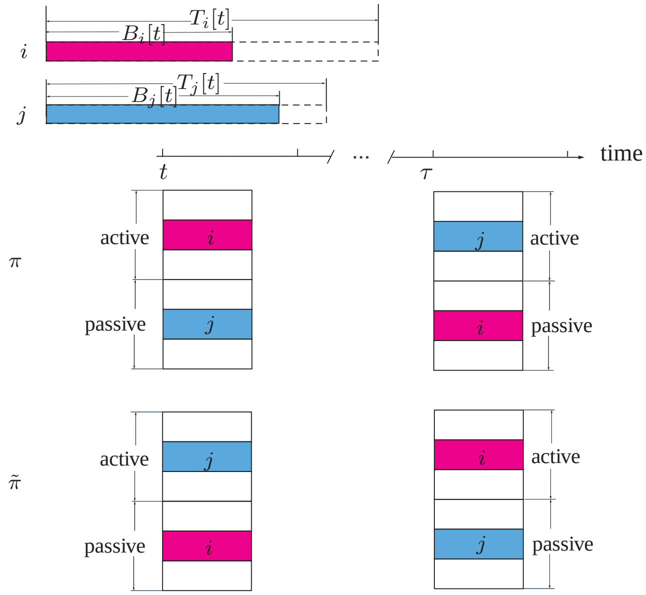

Fig. 3: The LLLP interchange

LLLP defines a partial order over the EVs' states such that the EV with less laxity and longer remaining charging demand should be given priority. In [5], the authors applied interchange argument to show that LLLP could improve the performance of any given policy along every sample path, and further, there exists an optimal stationary policy that follows the LLLP principle under mild conditions.

The LLLP interchange can be easily implemented to improve any given policy π. As illustrated in Fig. 3, suppose that at time t, EV j has less laxity and longer remaining charging demand than EV i ($j\succeq i$), and that the policy π charges i but not j. An LLLP interchange improved policy $\tilde{\pi}$ charges j but not i at time t. Let $\tau\in[t+1,\min\{d_i,d_j\}]$ denote the time period at which π charges j but not i for the first time; at time τ, $\tilde{\pi}$ charges i but not j. If such a period τ does not exist, then the interchanging policy $\tilde{\pi}$ will take the same action as the original policy π after time t.

The LLLP principle results in many smaller unfinished jobs rather than few large unfinished jobs and thus improves the profit performance under time-varying charging costs and convex non-completion penalty function. The LLLP principle will be used to improve the Whittle's index policy.

B. Index Policy with LLLP interchange

In this subsection, we propose a heuristic policy: the Whittle's index policy with LLLP interchange. The heuristic policy can be obtained by implementing Algorithm 1.

The proposed policy takes the advantage of time varying charging cost while balancing the risk of non-completion penalties. In principle, it gives higher priority to EVs with tight deadlines and large remaining demand to avoid potential penalties.

Algorithm 1 Whhittle Index with LLLP interchange

1. Calculate the Whittle's index of all chargers and sort them in a descend order.
2. Apply LLLP inter-change to the sorted chargers.
3. Activate the M chargers with highest priority.

V. Numerical results

In this section, numerical experiments are conducted to compare the performance of proposed index policy with other simple heuristic (index) policies, *i.e.*, EDF (earliest deadline first), LLF (least laxity first) and Whittle's index without LLLP interchange. If feasible, EDF charges M EVs with the earliest deadlines, and LLF charges M EVs with the least laxity. Both policies will fully utilize the capacity and charge M EVs as long as there are at least M unfinished EVs in the system. Whittle's index policy, on the other hand, ranks all chargers by the Whittle's index and activates the first M, and may put some (regular) chargers idle when the cost is high.

We first consider a special case of Problem (8) with a constant charging cost. Since the charging cost is time-invariant, it is optimal to fully utilize the charging capacity M to charge unfinished EVs.

We observe from Fig. 4 that the Whittle's index policy with LLLP interchange and LLF achieve similar performance, since both policies roughly follow the least laxity first principle. The performance of these two policies is close to the performance upper bound. The EDF policy performs poorly because it does not take the remaining charging demand into account. The gap between the Whittle's index policy and the index policy with LLLP interchange comes from the reordering of EVs with positive laxity (cf. the discussion following Theorem 1).

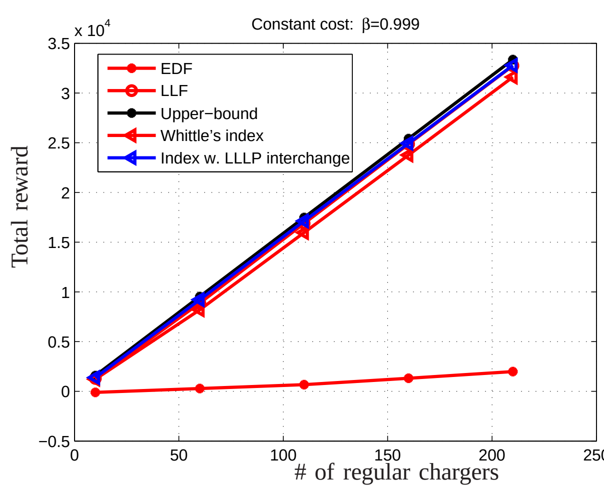

Fig. 4: Performance comparison: constant charging cost $c[t] = 0.5$, $\rho = 0.7$, $T_{\max} = 12$, $B_{\max} = 9$, $\beta = 0.999$, $M/N = 0.5$.

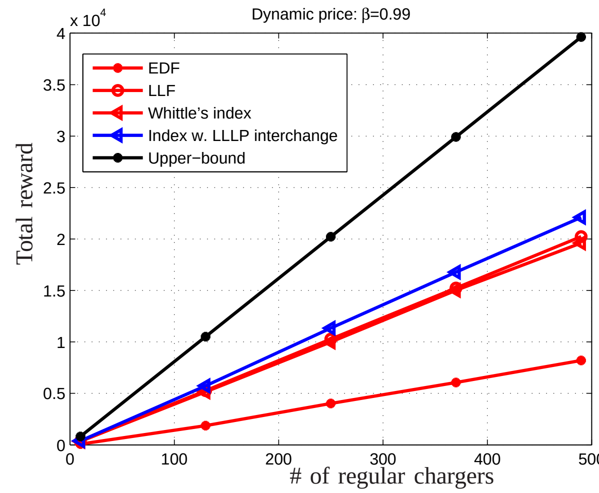

Fig. 5: Performance comparison: dynamic charging cost, $\rho = 0.7$, $T_{\max} = 12$, $B_{\max} = 9$, $\beta = 0.999$, $M/N = 0.5$.

For the dynamic charging cost case, we use the real-time pricing signal from NY Independent System Operator (NYISO) and train a Markovian model that describes marginal charging costs. The Markov chain of charging cost has a period of 24 hours (a single day). In Fig. 5, the performance of different policies is compared. Both EDF and LLF seeks to activiate as many regular chargers as possible, up to the capacity constraint M. The Whittle's index policy, on the other hand, takes the advantage of the pricing fluctuation and charges more EVs at price valley and keeps some chargers idle when the cost is too high. Based on the Whittle's index policy, LLLP can reduce the number of unfinished EVs with large remaining charging demand and therefore reduce the non-completion penalties. The total reward achieved by the Whittle's index with LLLP interchange policy is more than twice of that obtained by EDF; the performance gap between the the Whittle's index with LLLP interchange policy and the LLF policy is over 10%. We also note that the LLLP principle improves Whittle's index by over 10%.

VI. Conclusion

In this paper, we considered the problem of scheduling of the charging of a large number EVs in public facilities—a problem of particular potential significance as EV penetration deepens. In such settings, it is essential to develop highly efficient and online charging algorithms. To this end, index policies considered in this paper are attractive for its implementation simplicity and versatility in incorporating various operation uncertainties.

References

[1] Hybridcars, "2011-2015 plug-in vehicles monthly sales dashboard," 2015. Available from: http://www.hybridcars.com/august-2015-dashboard/.

[2] U.S. Department of Energy, “Electric vechicle charging station locations,” 2015. Available from: http://www.afdc.energy.gov/fuels/electricity_locations.html.

[3] Z. Yu, S. Li, and L. Tong, “On Market Dynamics of Electric Vehicle Diffusion,” in *Proceedings of the 52nd Annual Allerton Conferene on Communication, Control, and Computing*, Oct. 2014.

[4] P. Whittle, “Restless bandits: Activity allocation in a changing world,” *Journal of applied probability*, pp. 287–298, 1988.

[5] Y. Xu and F. Pan, “Scheduling for charging plug-in hybrid electric vehicles,” in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 2495–2501, IEEE, 2012.

[6] V. Raghunathan, V. Borkar, M. Cao, and P. R. Kumar, “Index policies for real-time multicast scheduling for wireless broadcast systems,” in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, IEEE, 2008.

[7] J. J. Jaramillo, R. Srikant, and L. Ying, “Scheduling for optimal rate allocation in ad hoc networks with heterogeneous delay constraints,” *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 5, pp. 979–987, 2011.

[8] S. S. Panwar, D. Towsley, and J. K. Wolf, “Optimal Scheduling Policies for a class of Queues with Customer Deadlines to the Beginning of Service,” *Journal of Association for Computing Machinery*, vol. 35, pp. 832–844, October.

[9] D. Towsley and S. Panwar, “On the Optimality of Minimum Laxity and Earliest Deadline Scheduling for Real-Time Multiprocessors,” in *Proc. of IEEE Euromicro 90’ Workshop on Real-Time*, pp. 17–24, Jun. 1990.

[10] K. R. L. Kruk, J. Lehoczky and S. Shreve, “Heavy traffic analysis for EDF queues with reneging,” *Annals of Applied Probability*, vol. 21, no. 2, pp. 484–545, 2011.

[11] C. L. Liu and J. W. Layland, “Scheduling algorithms for multiprogramming in a hard-real-time environment,” *Journal of ACM*, vol. 20, pp. 46–61, 1973.

[12] M. Dertouzos, “Control robotics: the procedural control of physical processes,” in *Proc. IFIP Congress*, pp. 807–813, 1974.

[13] A. Mok, *Fundamental design problmes of distributed systems for the hard real-time environment*. PhD thesis, MIT, 1983.

[14] M. L. Dertouzos and A. K. Mok, “Multiprocessor online scheduling of hard-real-time tasks,” *IEEE Transactions on Software Engineering*, vol. 5, pp. 1497–1506, 1989.

[15] S. Chen, L. Tong, and T. He, “Optimal deadline scheduling with commitment,” in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pp. 111–118, IEEE, 2011.

[16] S. Chen, Y. Ji, and L. Tong, “Large scale charging of electric vehicles,” in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–9, IEEE, 2012.

[17] S. Chen and L. Tong, “iems for large scale charging of electric vehicles: Architecture and optimal online scheduling,” in *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, pp. 629–634, IEEE, 2012.

[18] A. Subramanian, M. Garcia, A. Dominguez-Garcia, D. Callaway, K. Poolla, P. Varaiya, *et al.*, “Real-time scheduling of deferrable electric loads,” in *American Control Conference (ACC), 2012*, pp. 3643–3650, IEEE, 2012.

[19] L. Gan, U. Topcu, and S. H. Low, “Optimal decentralized protocol for electric vehicle charging,” *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 940–951, 2013.

[20] Z. Ma, D. S. Callaway, I. Hiskens, *et al.*, “Decentralized charging control of large populations of plug-in electric vehicles,” *Control Systems Technology, IEEE Transactions on*, vol. 21, no. 1, pp. 67–78, 2013.

[21] Q. Li, T. Cui, R. Negi, F. Franchetti, and M. D. Ilic, “On-line decentralized charging of plug-in electric vehicles in power systems,” *arXiv preprint arXiv:1106.5063*, 2011.

[22] C. Wu, H. Mohsenian-Rad, and J. Huang, “Vehicle-to-aggregator interaction game,” *Smart Grid, IEEE Transactions on*, vol. 3, no. 1, pp. 434–442, 2012.

[23] J. Donadee and M. D. Ilic, “Stochastic optimization of grid to vehicle frequency regulation capacity bids,” *Smart Grid, IEEE Transactions on*, vol. 5, no. 2, pp. 1061–1069, 2014.

[24] F. Juul, M. Negrete-Pincetic, J. MacDonald, and D. Callaway, “Real-time Scheduling of Electric Vehicles for Ancillary Services,” in *Proc. 2015 IEEE Power and Energy Society General Meeting*, (Denver, CO), 2015.

[25] E. Altman, *Constrained Markov decision processes*, vol. 7. CRC Press, 1999.

[26] J. Nino-Mora *et al.*, “Restless bandits, partial conservation laws and indexability,” *Advances in Applied Probability*, vol. 33, no. 1, pp. 76–98, 2001.