Demand Response via Large Scale Charging of Electric Vehicles

Zhe Yu† and Lang Tong†

***Abstract*—The problem of centralized scheduling of large scale charging of electric vehicles (EVs) with demand response options is considered. A stochastic dynamic programming model is introduced in which the EV charging service provider faces stochastic demand, convex non-completion penalties, and random demand response requirements. Formulated as a restless multi-armed bandit problem, the EV charging problem is shown to be indexable, thus low complexity index policies exist. 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. Numerical results illustrate the performance improvement and the capability of handling various operation uncertainties of the proposed index policy.**

***Index Terms*—Multi-armed bandit problem; Deadline scheduling; Charging of electric vehicles; Whittle's index; Demand response.**

I. Introduction

WITH the substantially growth of Electric Vehicles (EVs) and EV charging services [1], [2], the potential of participating in demand response programs by EV charging service providers has attracted considerable interest. To a EV charging service provider who has the capacity of serving a large number of EVs, the economic benefit of providing demand response can be substantial.

In this paper, we consider the problem of providing a form of ancillary service by a large scale EV charging service provider with fast charging capabilities and the capacity of serving hundreds of vehicles. An example of such a service provider can be one that operates at large public or private parking facilities. An essential characteristic of such charging services is that it can shift substantial demand without seriously jeopardizing the quality of service because, among the large number of EVs in the facility, there is a substantial laxity in fulfilling the charging demand.

However, 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, there is significant uncertainty in charging demand and charging cost. 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. The real-time electricity price may be fluctuating and the local renewables such as solar generations may be intermittent, which makes the charging cost random. Second, the aggregator needs to balance the charging demand and the demand response requirement. EV consumers desire their pre-declared state-of-charge (SOC) at departure which may be conflicting with the demand response profile. 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.

Z. Yu† and L. Tong† are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA. Email: {zy73,lt35}@cornell.edu.
This work is supported in part by the National Science Foundation under Grant CNS-1248079. Part of this work is presented at the 2015 Allerton conference in October 2015.

A. Related work

There is expanding literature on the EV charging with demand response. In [3] and [4], authors showed that single EV can be used to provide ancillary service and energy to the grid. Different from the centralized scheduling of EV charging framework considered here, distributed pricing strategy and algorithm are studied in [5] and [6] to encourage EVs to participate in frequency regulation. In [7] and [8], two-settlement central control algorithms are proposed. Charging trajectories of EVs are optimized day ahead and adjustment is carried out in real-time. In [9], authors investigated the real-time adjustment balancing the tracking of predetermined charging trajectories and regulation signal. However, a real-time algorithm that is scalable and robust to various uncertainties is lacking.

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 [10] by Raghunathan, Bokar, and Kumar on a deadline scheduling problem in wireless communications. The authors of [10] are perhaps the first to formulate the stochastic deadline scheduling problem as a restless MAB problem and established indexability. Also related is [11] where the problem of scheduling packets with deadlines in ad hoc networks is considered. There are several nontrivial differences between the models in [10], [11] and that in the current paper. For instance, the arrival models used in [10] are either simultaneous or periodic. The cost models in [10] and [11] are also significantly different from ours.

The dynamic programming approach to EV charging was considered in [12] 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.

This paper extends the results in [13], where large scale EV charging without demand response is considered.

B. Summary of results

We introduce a stochastic dynamic programming model for large scale EV charging which captures randomness in arrival-

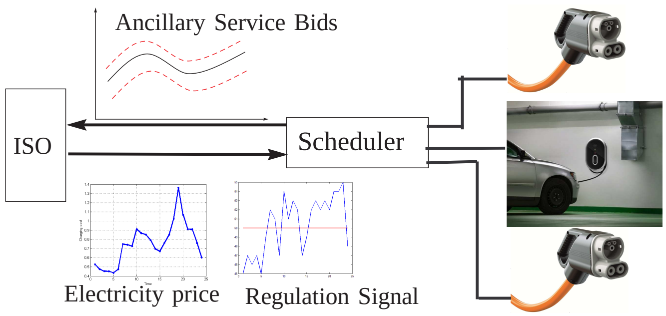

Fig. 1: Architecture of a charging station

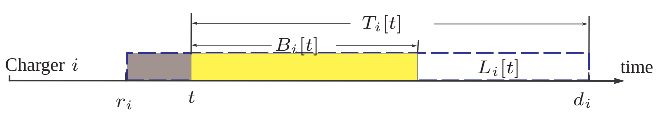

Fig. 2: An illustration for the charger's state. r_i is the arrival time of an EV at charger i, d_i the deadline for completion, $B_i[t]$ the amount of charging to be completed by d_i, $T_i[t]$ the lead time to deadline.

s, demand, deadlines, charging costs, and demand response requirement. To handle scalable solution, we first establish indexability of the problem and seek an index policies with computable indices. Numerical simulations demonstrate that the proposed policy makes significant charging profit while providing perfect ancillary service under various uncertainties.

II. Problem formulation

Fig. 1 shows a schematic of an energy management system at an EV charging facility. The charging facility has N parking spots, each with a charger that can be activated or deactivated by the scheduler. EVs arrive at chargers independently. At the arrival of charger i, if the charger is not occupied, 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.

To participate in the demand response market, the scheduler submits the regulation mid point and regulation capability to the Independent System Operator (ISO) day ahead. In the operation day, the ISO sends out the real-time electricity price and a regulation signal for the scheduler to track.

We summarize the assumptions in the paper as follow.

A1. Each charger can be connected to only one EV, and it is removed from the EV at the deadline d_i.

A2. An EV is charged at a fixed rate normalized to 1 and can not be discharged [14].

A3. The EV arrivals to the N chargers are independent and identically distributed (i.i.d.).

A4. The price of charging collected from consumers is proportion to the charging demand, normalized to 1 dollar/hour.

A5. 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.

A6. The charging of EVs is preemptive without cost.

A7. The penalty for incomplete charging is a convex function of the incomplete amount at the deadline.

A8. The regulation signal is stochastic and independent from the actions of the charging facility.

We now present elements of the discrete-time stochastic dynamic programming 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.

A. State space

The state of the charging system

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

is defined by the regulation signal $M[t]$, the charging cost $c[t]$, and states of individual chargers $S_i[t]$ where $\mathcal{M}$ is the state space of the regulation signal, $\mathcal{S}_c$ 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 charging cost $c[t]$ is the cost of electricity from the wholesale market, offset by possibly locally generated renewables. The regulation signal $M[t]$ is assumed to be integers.

B. Action and State evolution

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).

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 ($B_i[t] = 0$), 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

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

where ρ is the probability of EV arrivals.

The state transition under the passive action is similar, except that only the lead time is decreasing.

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

The charging cost $c[t] \sim (\mathcal{S}_c, P)$ is assumed as an exogenous finite state Markov chain, independent of the actions of the scheduler and individual charger state evolutions, with transition probability matrix $P = [P_{i,j}]$.

C. Charging Profit

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.

The charging profit collected from the EVs is stated as

$$\mathcal{R}_{\text{Ch}}(S[t], \vec{a}[t]) \triangleq \sum_{i=1}^{N} R_{a_i[t]}(S_i[t], c[t]).$$

D. Demand Response Credit

Based on the new rule of demand response market [15], the credit includes the capacity payment and performance payment. The former one is simply the regulation amount times the regulation price per unit. The performance payment is measured by the tracking accuracy of the regulation signal. The accuracy as defined in PJM ancillary service market is stated as $(1 - |\sum_{i=1}^{N} a_i[t] - M[t]|/M[t])$ [16], where $M[t]$ is the regulation signal and $\sum_{i=1}^{N} a_i[t]$ the extra power usage.

The demand response credit collected from the ISO at time t is stated as follows.

$$\mathcal{R}_{\text{AS}}(S[t], \vec{a}[t]) \triangleq A[t](1 - |\sum_{i=1}^{N} a_i[t] - M[t]|/M[t]) + B[t],$$

where $A[t]$ and $B[t]$ take account of the stochastic demand response price and capacity credit.

E. Objective and optimal policy

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

$$V_\pi(s) \triangleq \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \beta^t [\mathcal{R}_{\text{Ch}}(S[t], \vec{a}[t]) + \mathcal{R}_{\text{AS}}(S[t], \vec{a}[t])] \middle| S[0] = s \right\} \tag{4}$$

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

The optimal policy now can be formulated as follows.

$$V(s) = \sup_\pi V_\pi(s), \tag{5}$$

A policy π^* is said optimal if $V_{\pi^*}(s) = V(s)$.

III. Index policy and Whittle's index

The stochastic dynamic programming formulation does not result in a scalable scheduling policy. The problem complexity is exponential in the number of chargers and the randomness in regulation signal and price introduces extra difficulties.

Since 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 an 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.

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.

Since the cost dynamic is independent of the state and actions of chargers, the state transition of arm i can be written according to charger transition (1), (2) and cost transition P.

2) MAB formulation: In the traditional MAB problem, the arms are independent and the objective is to maximize the sum of the rewards collected from each arm. However, the objective of (5) includes the demand response credit which couples arms. The random regulation signal and regulation price introduce extra complexity. One intuitive way is to model the tracking of the signal as a constraint on number of active arms and to maximize the charging rewards. The stochastic dynamic programming in (5) can be viewed as a restless MAB problem that, at each time t, exactly $M[t]$ out of N chargers (arms) are active. The optimization problem is state as following:

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

B. Whittle's index

We now examine the Whittle's index policy for the restless MAB problem defined in (6). 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 $\tilde{S}_i[0]$,

$$V_i(\tilde{s}) \triangleq \sup_\pi \mathbb{E}_\pi \left\{ \sum_{t=0}^{\infty} \beta^t R_{a_i[t]}(\tilde{S}_i[t]) \mid \tilde{S}_i[0] = \tilde{s} \right\}, \tag{7}$$

where V_i is the value function. Let $\mathcal{L}_a$ be the Markov transition operator on 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 (7) 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 [17]. 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 the following theorem, the indexability of the MAB problem in (6) is established and the closed-form expression of Whittle's index is derived for the case with constant charging cost. For dynamic cost case, the Whittle's index can be obtained by solving a parametric programming [18].

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 is given by

$$\nu_i(T, B, c_0) = \begin{cases} 0 & \text{if } B = 0, \\ 1 - c_0 & \text{if } 1 \leq B \leq T-1, \\ 1 - c_0 + \beta^{T-1}[F(B-T+1) - F(B-T)] & \\ & \text{if } T \leq B. \end{cases} \quad (8)$$

The proof of Theorem 1 is omitted due to the space limit.

In (8), 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.

IV. Whittle's index policy with LLLP interchange

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.

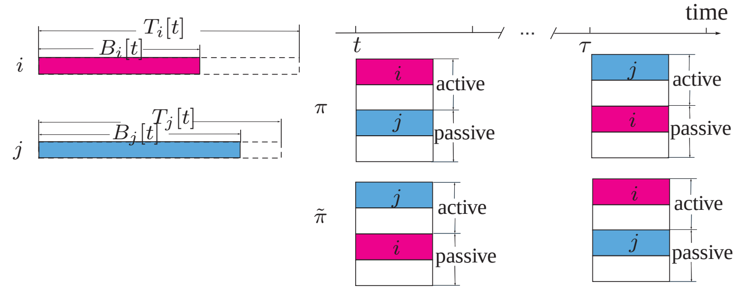

Fig. 3: The LLLP interchange

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.*, $B_j[t] \geq B_i[t]$ and $L_j[t] \leq L_j[t]$, with at least one of the inequalities being strict.

LLLP defines a partial order over the EVs' states. In [12], 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.

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 tries the best to follow the regulation signal and 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 EVs and sort them in a descend order.
2. Apply LLLP inter-change to the sorted EVs.
3. Activate the $M[t]$ EVs with highest priority.

Note that the proposed algorithm does not guarantee the feasibility. If there is not enough EVs with positive charging demand in the system, we simply charge as many as possible.

V. Numerical results

In this section, numerical experiments are conducted to compare the performance of different scheduling policies with demand response options. One intuitive trajectory tracking algorithm is proposed in [9], which minimizes a trade off between the tracking errors of a predetermined charging trajectory for each vehicle and the deviation from the regulation signal. In the simulation, we first apply Algorithm (1) with $M[t] = M$ to generate a charging trajectory for each charger. After that, the regulation signal is generated by a uniform distribution with mean M. Since the result of the convex programming in [9] is continuous, the binary charging action is generated by Bernoulli random variables with the results of the convex programming as the probability coefficients.

Fig. 4 shows an example of regulation signal tracking performance of different policies. When the constraint is feasible, the Whittle's index policy with LLLP interchange always perfectly matches the regulation signal. While, the trajectory tracking policy deviates from the regulation signal slightly due to the difference between the predetermined charging trajectory and the regulation signal.

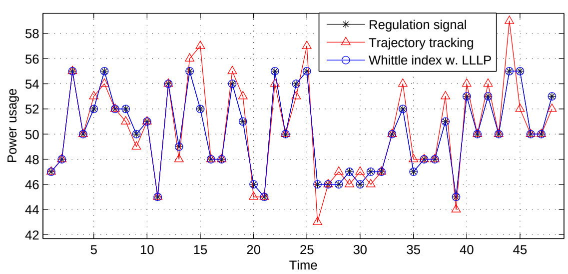

Fig. 4: Regulation signal tracking: $M = 50$, 10% regulation capacity.

The impact of the regulation capacity on tracking accuracy and charging rewards is illustrated in Fig. 5. As the capacity increases, the tracking error of trajectory tracking policy increases while Whittle Index policy with and without LLLP improvement matches the regulation signal perfectly. LLLP improves the Whittle's index policy in the charging reward collected from EV consumers and thus makes more profit.

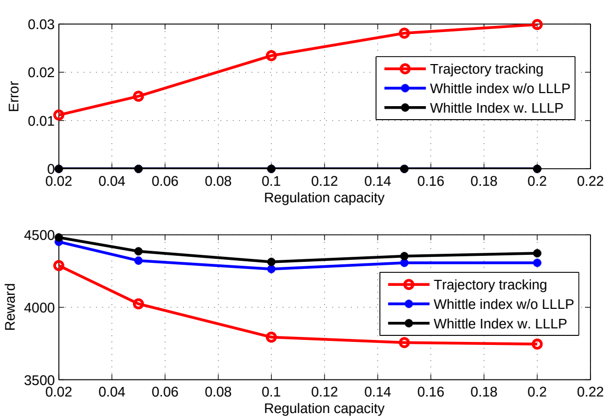

Fig. 5: Regulation signal tracking accuracy and charging reward Vs. regulation capacity.

VI. Conclusion

In this paper, we considered the demand response in the scheduling of the charging of a large number EVs in public facilities, which will significantly benefit both the grid and the EV customers as EV penetration deepens. Due to the curse of the dimension, it is essential to develop a highly efficient and scalable online algorithm. Index policies considered in this work show the implementation simplicity and capability to handle 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] W. Kempton and J. Tomić, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.

[4] A. N. Brooks, *Vehicle-to-grid demonstration project: Grid regulation ancillary service with a battery electric vehicle*. California Environmental Protection Agency, Air Resources Board, Research Division, 2002.

[5] 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.

[6] 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.

[7] S. Vagropoulos, A. G. Bakirtzis, *et al.*, "Optimal bidding strategy for electric vehicle aggregators in electricity markets," *Power Systems, IEEE Transactions on*, vol. 28, no. 4, pp. 4031–4041, 2013.

[8] E. Sortomme, M. El-Sharkawi, *et al.*, "Optimal scheduling of vehicle-to-grid energy and ancillary services," *Smart Grid, IEEE Transactions on*, vol. 3, no. 1, pp. 351–359, 2012.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Z. Yu, Y. Xu, and L. Tong, "Large Scale Chgarging of Electric Vehicles: a Multi-Armed Bandit Approach," in *Proceedings of the 52nd Annual Allerton Conferene on Communication, Control, and Computing*, Oct. 2015.

[14] S. Han, S. Han, and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 65–72, 2010.

[15] PJM, "Pjm manual 28: Operating agreement accounting," 2015. Available from: http://www.pjm.com/~/media/documents/manuals/m28.ashx.

[16] PJM, "Pjm manual 12: Balancing operations," 2015. Available from: http://www.pjm.com/~/media/documents/manuals/m12.ashx.

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

[18] J. Niño-Mora, "Characterization and computation of restless bandit marginal productivity indices," in *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, p. 74, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.