

# Robust Online Algorithms for Peak-Minimizing EV Charging under Multi-Stage Uncertainty

Shizhen Zhao and Xiaojun Lin

School of ECE,

Purdue University, West Lafayette, IN, USA

Email: {zhao147, linx}@purdue.edu

Minghua Chen

Department of Information Engineering,

The Chinese University of Hong Kong

Email: minghua@ie.cuhk.edu.hk

**Abstract**—In this paper, we study how to utilize forecasts to design online EV (electrical vehicle) charging algorithms that can attain strong performance guarantees. We consider the scenario of an aggregator serving a large number of EVs together with its background load, using both its own renewable energy (for free) and the energy procured from the external grid. The goal of the aggregator is to minimize its peak procurement from the grid, subject to the constraint that each EV has to be fully charged before its deadline. Further, the aggregator can predict the future demand and the renewable energy supply with some levels of uncertainty. We show that such prediction can be very effective in reducing the competitive ratios of online control algorithms, and even allow online algorithms to achieve close-to-offline-optimal peak. Specifically, we first propose a 2-level increasing precision model (2-IPM), to model forecasts with different levels of accuracy. We then develop a powerful computational approach that can compute the optimal competitive ratio under 2-IPM over any online algorithm, and also online algorithms that can achieve the optimal competitive ratio. Simulation results show that, even with up to 20% day-ahead prediction errors, our online algorithms still achieve competitive ratios fairly close to 1, which are much better than the classic results in the literature with a competitive ratio of  $e$ . The second contribution of this paper is that we solve a dilemma for online algorithm design, e.g., an online algorithm with good competitive ratio may exhibit poor average-case performance. We propose a new *Algorithm-Robustification* procedure that can convert an online algorithm with good average-case performance to one with both the optimal competitive ratio and good average-case performance. The robustified version of a well-known heuristic algorithm based on Model Predictive Control (MPC) is found to demonstrate superior performance via trace-based simulations.

## I. INTRODUCTION

Replacing fossil fuels by renewable energy is a major priority all over the world (see, e.g., [2][3] and the references therein). However, high penetration of renewable energy challenges the reliable and efficient operation of the power grid. Specifically, renewable energy from wind and solar is known to exhibit high variability and uncertainty. As renewable generation varies, the grid needs additional flexibility to balance the demand and supply [4]. In this paper, we focus on balancing the variability and uncertainty of the renewable supply by exploiting the flexibility from electric vehicle (EV) charging

demand [5], which is a typical example of deferrable demands [6]. As reported in [7], the sale of EVs is expected to increase exponentially, and 50% of new cars will be EVs by 2040. Thus, the future EV demand can potentially be huge, which could be used to compensate the variability and uncertainty due to high penetration of renewable energy.

Towards this end, in this paper we study how to develop robust online EV charging algorithms that minimize the impact of variability and uncertainty of renewable energy to the grid. Specifically, we consider a demand-aggregator who has its own background demand and renewable energy supply (the latter is assumed to be of no cost), and who manages a large number of EVs. Such an aggregator could represent an apartment or office building with a parking garage, a campus, or a micro-grid [8]. As the EVs arrive and are connected to the charging stations, each of them specifies a deadline for the charging request to be completed. The aggregator will first use its own renewable energy to serve its background demand and EV demand, and if that is not enough, it will purchase additional energy from the grid. We model the objective of the aggregator as minimizing the peak consumption from the grid under the constraints that all EVs must be charged before their deadlines. Our choice of the peak-minimization objective is motivated by the following two considerations. First, a large peak consumption-level requires the grid to provision the corresponding generation and transmission capacity in order to meet the demand. Thus, a large peak not only increases the overall cost of supplying energy, but also poses danger to grid-stability. Second, utility companies have already developed peak-based pricing schemes to encourage large customers (including aggregators) to reduce their peak and smoothen their demand. In this type of pricing schemes, the customers are charged based on not only the total usage in a billing period, but also the maximum (peak) usage at any time in the billing period. Specifically, if a customer's energy consumption is given as a sequence  $(E_1, E_2, \dots, E_n)$ , then the total bill is of the form  $c_1 \sum_i E_i + c_2 \max_i \{E_i\}$  [9]. As one can see from an example of such a peak-based pricing scheme from the National Grid [9], the average charge for peak usage  $c_2$  (8.32\$/kW-month) is over 100 times more than the unit charge for total usage  $c_1$  (0.07\$/kWh). As a result, the peak charge can constitute a large fraction (ranging from 30% to 50% [10]) of the total electricity bill. Under this type of pricing schemes, when the aggregator reschedules EV charging jobs, the total energy consumption from the grid does

An earlier version of this work has been presented at IEEE INFOCOM 2015 [1].

not change. It is the peak demand that is changed. Hence, minimizing the aggregator's operating cost is also equivalent to minimizing its peak consumption. Further, the potential benefit of peak reduction is huge. For campus-level aggregators (e.g., [8]), the peak energy is usually on the order of 20MW. Then, every one percent of peak reduction will correspond to  $0.01 \times 20\text{MW} \times \$9/\text{kW-month} \times 12 = \$21600$  saving per year.

The main difficulty in the above EV charging problem comes from the *sequentially-revealed* uncertainty in both the demand and the renewable supply. If all the demand and the supply could be precisely predicted in advance, one could have used an offline algorithm to compute the optimal charging schedule that minimizes the peak. Specifically, there exist both centralized algorithms (e.g., the YDS algorithm in [11]) and decentralized algorithms (e.g., the valley-filling algorithm in [12]) for peak-minimizing EV-charging in an offline setting. However, in reality both the future demand and supply can exhibit significant uncertainty. Further, such uncertainty is typically sequentially revealed. That is, at any point in time, the past demand and supply are revealed, the future uncertainty is still unknown, yet the aggregator must make decision right away. Model Predictive Control (MPC) can be used to modify an offline algorithm to deal with sequentially-revealed uncertainty [13]. However, as readers will see in the example in Section II-C, an MPC-based algorithm could lead to much larger peak consumption levels (a similar observation was also reported in [12]).

There are various approaches in the literature to deal with control problems with uncertainty. However, as we elaborate below, these existing approaches are not suitable for the peak-minimizing EV-charging problem that we study in this paper. If a probabilistic model is known for future uncertainty, then the problem can be cast as a stochastic control problem. However, obtaining an accurate probabilistic model of uncertainty can be challenging, especially when the renewable supply is non-stationary and highly-correlated across time. Further, the complexity of solving the optimal control decision for a given probabilistic distribution, e.g., using Markov Decision Processes [14], is extremely high. Another way to deal with uncertainty is robust optimization [15][16], where future uncertainty is modeled in a set. The resulting solution is designed to optimize the worst-case performance for all possible realizations of the uncertainty in this set. However, robust optimization typically does not deal with sequential decisions.

The third approach, which we adopt in this work, is to design competitive online algorithms [17]. In the computer science literature, online algorithms are specifically designed to deal with the case where the decisions must be made sequentially based on uncertain input that is sequentially revealed. In a peak-minimizing problem closely related to ours [18], it was shown that, even without any future information of job arrivals and deadlines, one can design a competitive online algorithm whose peak consumption is at most a constant factor  $e = 2.718$  above the offline optimal (where the latter assumes that the future information is known in advance). This constant factor is referred to as the competitive ratio of the online algorithm. However, this line of research also encounters a number of challenges. First, existing results on competitive

online algorithms either are based on very simple models of future uncertainty [19], or do not assume any model at all. As a result, the worst-case performance and the corresponding competitive ratio are often quite poor. In practice, both renewable supply and EV demands can be predicted to a certain degree. Intuitively, such prediction can provide very useful information for eliminating uninteresting worst cases, and thus sharpening the competitive ratio of online algorithms. However, to the best of our knowledge, there is no systematic methodologies for designing competitive online algorithms under more sophisticated models of future uncertainty. The second challenge, which in fact applies to typical robust-optimization results as well [16], is that the algorithms are only optimized for the worst-case. As a result, their average-case performance can be quite poor [19]. Given that the worst-case input may only occur very rarely, the aggregator may then be hesitant to endorse the resulting algorithm.

In this work, we make two contributions that precisely address these difficulties. First, from the *methodology* point of view, we extend the framework of online algorithm to incorporate available future knowledge captured by prediction. Specifically, we propose a more general set-based model, called 2-IPM (2-level increasing precision model), to capture the *sequentially* revealed uncertainty of renewable supply, EV demand, and background load. Compared to the traditional set-based model used in robust optimization [15], a key novelty of 2-IPM is that it can model the *sequential* nature of multiple predictions, i.e., predictions can be made at multiple instants (e.g., day-ahead prediction versus intra-day prediction), and the predictions closer to the target time tend to be more accurate (e.g., intra-day prediction is usually more accurate than day-ahead prediction). For any given 2-IPM model, we then develop a powerful computation procedure to find the smallest competitive ratio in terms of the peak consumption. This smallest competitive ratio can thus be viewed as a measure of the “price of uncertainty” under the 2-IPM. As readers will see in Section V-B, our 2-IPM yields much lower price-of-uncertainty compared to the uncertainty models in [19].

Second, from the *algorithm* point of view, we propose a general “robustification” procedure to design online algorithms to address both worst-case performance and average-case performance. Specifically, given any online algorithm with good average performance (in terms of the peak), this robustification procedure can convert it to one with not only good average-case performance, but also the optimal competitive ratio. We apply this robustification procedure to a well-known online algorithm, called Shrinking Horizon Model Predictive Control (SH-MPC), which demonstrates good average-case performance, but poor worst-case competitive ratio. Our numerical results in Section V-C indicates that the robustified-SH-MPC algorithm achieves both good average-case and worst-case performance.

The rest of the paper is organized as follows. Section II defines the system model that captures the renewable uncertainty, and motivates the design of online algorithms with the optimal competitive ratio. We derive a fundamental lower bound on the competitive ratio in Section III, and propose

a general framework based on Algorithm Robustification in Section IV to design online algorithms that achieve the above lower bound. Real-trace based simulation results are provided in Section V to demonstrate both the improvement on the optimal competitive ratio and the effectiveness of our Algorithm Robustification procedure. In Section VI, we conclude.

## II. SYSTEM MODEL

We consider an aggregator serving its EV demand and background demand using both its own renewable energy (which is assumed to be cost-free) and the energy procured from the external grid. We assume that time is slotted, and index a time-slot by an integer in  $\mathbb{T} = \{1, \dots, T\}$ , where  $T$  is the time-horizon considered. We represent the EV demand by a  $T \times T$  upper-triangular matrix  $\mathbf{a} = [a_{i,j}]$ , where  $a_{i,j}$  is the total deferrable (EV) demand with arrival time  $i$  and deadline  $j \geq i$ . We represent the net non-deferrable demand by a  $T \times 1$  vector  $\mathbf{b} = [b_i]$ , where  $b_i$  is the background demand at time  $i$  minus the renewable energy available at time  $i$ . Note that when the penetration of renewable energy is high, the net non-deferrable demand will exhibit significant uncertainty. Using the flexibility in the EV demand, the goal of the aggregator is to schedule EV charging jobs against high renewable uncertainty such that the peak energy procured from the grid is minimized.

### A. Model for Prediction and Uncertainty

In practice, there exists considerable uncertainty in both the net non-deferrable demand and the deferrable demand. Specifically, we define a  $(T - t + 2) \times 1$  vector  $\mathbf{x}(t) = [a_{t,t}, \dots, a_{t,T}, b_t]^T$  to include both the EV demand with arrival time  $t$  and the net non-deferrable demand at time  $t$ . Note that the aggregator will know the precise value of  $\mathbf{x}(t)$  only at and after time-slot  $t$ . In the rest of this paper, we will say that “the value of  $\mathbf{x}(t)$  is revealed at time  $t$ ”. At a time  $s < t$ , the value of  $\mathbf{x}(t)$  is uncertain to the aggregator. However, the aggregator can use various sources of information (such as weather forecast) to predict the future values of these uncertain quantities in order to improve its decision. In practice, such predictions can be taken multiple times, e.g., if the operating time-horizon is a day, one prediction can be made before the day (called “day-ahead” prediction), and another prediction can be made a few hours before time  $t$  (called “intra-day” prediction). In general, intra-day prediction is more accurate than the day-ahead prediction because it is closer to the real time. Next, we will present a model, called 2-IPM (2-Level Increasing Precision Model), to model the uncertainty associated with such prediction procedures. We note that, although for ease of exposition the model below only assume one intra-day prediction, both 2-IPM and the subsequent results can be easily generalized to multiple intra-day predictions.

Specifically, we assume that at time 0 (before the first time-slot), a day-ahead prediction is available for every  $\mathbf{x}(t)$ ,  $t \in \mathbb{T}$ . For each future time-slot  $t$ , the day-ahead prediction provides two  $(T - t + 2) \times 1$  vectors  $\hat{\mathbf{x}}^L(0, t)$ ,  $\hat{\mathbf{x}}^U(0, t)$ , which are lower and upper bounds, respectively, to  $\mathbf{x}(t)$  (see Fig. 1 (a)). In

other words, the future value of  $\mathbf{x}(t)$  must lie within

$$\hat{\mathbf{x}}^L(0, t) \leq \mathbf{x}(t) \leq \hat{\mathbf{x}}^U(0, t). \quad (1)$$

Then, at a later time  $u_t$ ,  $1 \leq u_t < t$ , another intra-day prediction is performed. (One example of  $u_t$  could be  $u_t = \max\{1, t - L\}$ , i.e., the intra-day prediction is performed  $L$  time-slots ahead.) The intra-day prediction provides another two  $(T - t + 2) \times 1$  vectors  $\hat{\mathbf{x}}^L(u_t, t)$ ,  $\hat{\mathbf{x}}^U(u_t, t)$ , that are *better* lower and upper bounds to  $\mathbf{x}(t)$  than the day-ahead prediction (see Fig. 1 (b)). In other words, the following will hold:

$$\hat{\mathbf{x}}^L(0, t) \leq \hat{\mathbf{x}}^L(u_t, t) \leq \mathbf{x}(t) \leq \hat{\mathbf{x}}^U(u_t, t) \leq \hat{\mathbf{x}}^U(0, t). \quad (2)$$

Obviously, a key difference between the day-ahead prediction and the intra-day prediction is that they are performed at different times. Thus, while the value of day-ahead prediction,  $\hat{\mathbf{x}}^L(0, t)$ ,  $\hat{\mathbf{x}}^U(0, t)$  for all  $t$ , are known even before time-slot 1, the value of  $\hat{\mathbf{x}}^L(u_t, t)$  and  $\hat{\mathbf{x}}^U(u_t, t)$  will not be known until time-slot  $u_t$ . (We will say that the value of  $\hat{\mathbf{x}}^L(u_t, t)$  and  $\hat{\mathbf{x}}^U(u_t, t)$  are revealed at time  $u_t$ .) However, from time-slot 0 to time-slot  $u_t - 1$ , although the aggregator does not know the future intra-day prediction for  $\mathbf{x}(t)$  that will be performed at time  $u_t$ , it does know that this future intra-day prediction will be more accurate. In order to model this knowledge, we assume that there exists a  $(T - t + 2) \times 1$  vector  $W(u_t, t) \leq \hat{\mathbf{x}}^U(0, t) - \hat{\mathbf{x}}^L(0, t)$ , which is known at time 0, that bounds the (future) intra-day prediction gap  $\hat{\mathbf{x}}^U(u_t, t) - \hat{\mathbf{x}}^L(u_t, t)$  (see Fig. 1 (b)), i.e.,

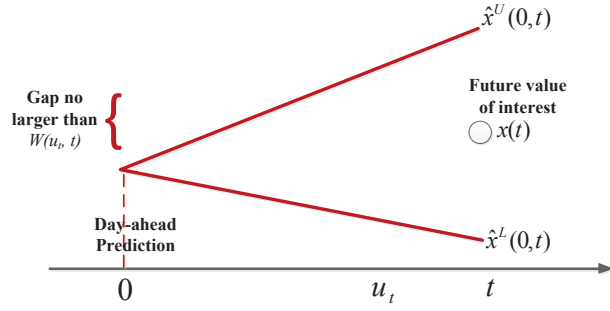
$$\hat{\mathbf{x}}^U(u_t, t) - \hat{\mathbf{x}}^L(u_t, t) \leq W(u_t, t). \quad (3)$$

In other words, the aggregator knows the (increased) precision level of future intra-day predictions that will be performed at time  $u_t$ , even though it does not know the exact bounds of this intra-day prediction before time  $u_t$ .

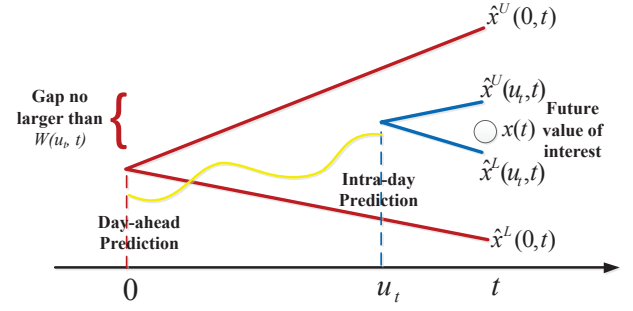
*Remark 1:* The novelty of our 2-IPM is in modeling the increasing precision of interval predictions. Note that compared to point predictions [20], interval predictions (i.e., (1) and (2)) provide additional information on the accuracy level of the prediction. While there have been many recent studies of interval predictions [21][22], the 2-IPM allows us to capture the increasing precision as time evolves and to rigorously study its effect on the corresponding sequential decision problem.

*Remark 2:* Readers may question what happens when the real value  $\mathbf{x}(t)$  or the bounds for the intra-day prediction falls outside of the day-ahead predicted interval (1). This is indeed possible in practice, because some predictions may be wrong. Nevertheless, for ease of theoretical analysis, we will assume that (1) and (2) are always satisfied. This assumption is justified because in practice these bounds are usually chosen such that the value of  $\mathbf{x}(t)$  will fall into the predicted intervals with high probability, and we can always tighten the intra-day prediction bounds so that (2) is satisfied. The case where these bounds are violated will be studied in Section IV-C and Section V-D.

We now define a few vector notations that summarize how the variables defined above are revealed in time. At time 0, the aggregator only knows  $Y = [\hat{\mathbf{x}}^L(0, t), \hat{\mathbf{x}}^U(0, t), W(u_t, t), t = 1, 2, \dots, T]$  (see Fig. 1 (a)). At the end of the time horizon  $T$ ,



(a) For any future quantity  $x(t)$  of interest, the day-ahead prediction and the width of the future intra-day prediction gap are known at time 0. We denote all such information by  $Y$ .



(b) The true value of  $x(u_t)$  and the intra-day prediction of  $x(t)$  are revealed at time  $u_t$ . We denote all the information revealed from time 1 to  $u_t$  by  $Z_{u_t}$ . We use  $Z$  to denote all information that will be revealed at the end of the time horizon  $T$ .

Fig. 1. Illustration of day-ahead prediction and intra-day prediction.

the aggregator knows not only  $Y$ , but also all revealed EV demand, net non-deferrable demand, and intra-day forecasts  $Z = [x(s), s = 1, 2, \dots, T, \hat{x}^L(u_s, s), \hat{x}^U(u_s, s), u_s \leq T]$ . Thus, we refer to this vector  $Z$  as a realization, which will only be revealed at the end of the time horizon. For any realization  $Z$ , let  $Z_t$  denote the sub-vector containing only those components revealed at or before time  $t$ , i.e.,  $Z_t = [x(s), s = 1, 2, \dots, t, \hat{x}^L(u_s, s), \hat{x}^U(u_s, s), u_s \leq t]$ . Note that the dimension of this sub-vector  $Z_t$  increases with time  $t$ . Similarly, we use  $Z_{>t}$  to denote the sub-vector of  $Z$  containing all components that are revealed after time  $t$ . Thus, at time  $t$ , the aggregator knows both  $Y$  and  $Z_t$ , but not  $Z_{>t}$ . Clearly, the set of components of  $[Z_t, Z_{>t}]$  is the same as that of  $Z$ . However, the order of the components of  $[Z_t, Z_{>t}]$  can be different from that of  $Z$ . Therefore, we write  $Z = \Pi_t([Z_t, Z_{>t}])$ , where  $\Pi_t$  is an appropriate one-to-one mapping that maps the components of  $[Z_t, Z_{>t}]$  to the corresponding components in  $Z$ .

Throughout this paper, we will view  $Y$  as given, because  $Y$  is known day-ahead (before any scheduling decisions are made). The uncertainty comes entirely from the realization  $Z$ . Even though we do not know the exact realization of  $Z$  beforehand, the knowledge of  $Y$  (the day-ahead prediction) restricts  $Z$  into a smaller sample space. Specifically, for a given  $Y$ ,  $Y$  and a possible realization  $Z$  must satisfy constraints (1)-(3), and all such possible  $Z$ 's form a demand-trace set  $\mathcal{Z}_Y$ , i.e.,

$$\mathcal{Z}_Y = \{Z : Z = [x(s), s = 1, 2, \dots, T, \hat{x}^L(u_s, s), \hat{x}^U(u_s, s), u_s \leq T], \text{ the components of } Z \text{ satisfy (1) - (3) for the fixed } Y\}.$$

In this paper, we aim to design policies that can attain strong performance guarantees for all possible realization drawn from  $\mathcal{Z}_Y$ .

### B. Objective

We are interested in designing online algorithms for scheduling EV demand that minimize the peak energy drawn from the grid. At each time  $t = 1, 2, \dots, T$ , an online algorithm  $\pi$  must determine the amount of energy  $E_t(Z_t, \pi)$  drawn from

the grid, *based only on the knowledge of  $Y$  and  $Z_t$* . (Note that we have assumed that  $Y$  is fixed. Hence, we have omitted  $Y$  in the notation  $E_t(Z_t, \pi)$  for simplicity, but the dependency of  $E_t(Z_t, \pi)$  on  $Y$  is implicitly assumed.) In other words, the decision at time  $t$  cannot be based on the values of any component of  $Z_{>t}$ , which will only be revealed in the future. The online algorithm  $\pi$  is said to be feasible if all the EV demands can be completely served before deadlines using the sequence of energy-procurement decisions  $[E_t(Z_t, \pi), t \in \mathbb{T}]$  minus the revealed non-deferrable demand (i.e., background demand minus renewable energy). Let  $E_\pi^p(Z) = \max_t \{E_t(Z_t, \pi)\}$  be the peak energy drawn from the grid using a feasible online algorithm  $\pi$ . The aggregator is interested on reducing  $E_\pi^p(Z)$ . However, it is not possible for one online algorithm to minimize  $E_\pi^p(Z)$  for all  $Z$ 's. Instead, we consider an *offline* solution provided by a "genie" that knows the entire future  $Z$  in advance. This genie can set the energy procurement  $E_t(Z)$  at each time-slot  $t$  based on  $Z$ . This genie can then solve the following problem offline:

$$\min_{\text{All demand can be completely served}} \max_t \{E_t(Z)\}. \quad (4)$$

Let  $E_{\text{off}}^*(Z)$  be the optimal offline solution to (4). Clearly, for any online algorithm  $\pi$ , we will have  $E_{\text{off}}^*(Z) \leq E_\pi^p(Z)$ . We can then evaluate the performance of an online algorithm  $\pi$  by comparing it to the above offline optimal. Specifically, for a fixed  $Y$ , define the competitive ratio (CR)  $\eta_Y(\pi)$  of an online algorithm  $\pi$  as the maximum ratio between  $E_\pi^p(Z)$  and  $E_{\text{off}}^*(Z)$  under all possible  $Z \in \mathcal{Z}_Y$ , i.e.,  $\eta_Y(\pi) = \max_{Z \in \mathcal{Z}_Y} \left\{ \frac{E_\pi^p(Z)}{E_{\text{off}}^*(Z)} \right\}$ .

In other words, the competitive ratio characterizes how *in the worst case* the online algorithm can perform more poorly compared to the offline optimal.

In the rest of the paper, we will first find an achievable lower bound on the competitive ratio  $\eta_Y(\pi)$  under 2-IPM, which characterizes the fundamental limits how 2-level prediction can improve the worst-case performance. Then, we will propose a systematic approach to design online algorithms with both the optimal competitive ratio and good average-case performance.

*Remark 3:* Related to the above peak-minimization objective, another approach is to set an upper limit for the peak

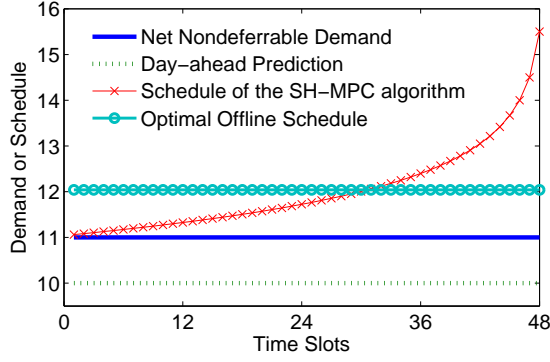


Fig. 2. Poor performance of the SH-MPC algorithm. In this example, the day-ahead prediction consistently underestimates the real demand, and the SH-MPC algorithm leads to a peak that is 1.29 times higher than the optimal offline peak. In comparison, an optimal online algorithm can achieve a peak that is at most 1.06 times higher than the optimal offline peak.

power consumption, and schedule EV charging within the power limit [23]. However, this approach does not minimize peak demand charges.

### C. A Motivating Example

Before describing our main results, we use an example to illustrate that it is not trivial to design an online algorithm with good competitive ratio.

Consider Model Predictive Control (MPC) [13], which is a popular approach for dealing with sequential-revealed uncertainty. At each time-slot, using the offline solution, MPC computes the entire schedule for the future based on past information that has been revealed and most-recently predicted future demand/supply. However, MPC only executes the first step of the schedule in the current time-slot. Then, in the next time-slot, as new information is revealed, MPC repeats this procedure with the newly-revealed information and again applies the first step. Note that, since our model assumes a finite time-horizon, the time window considered by MPC will shrink by one at each time-slot. Thus, this version of MPC is referred to as Shrinking-Horizon MPC (SH-MPC) in the literature [24]. Clearly, if the future demand/supply is close to their predicted values (which likely holds in the average case), the performance of SH-MPC will be quite good. However, the following example illustrates that the worst-case performance of SH-MPC can be very poor, especially when the future prediction is persistently wrong. Specifically, there are 48 time-slots. One EV charging job arrives at the beginning of time-slot 1, departs at the end of time-slot 48, and the total demand is 48. (In other words, there is no uncertainty for the EV demand.) The day-ahead predictions of the net non-deferrable demands are 10 for all time-slots, with an uncertain interval of  $[8, 12]$ , i.e., future non-deferrable demands are assumed to fall within this interval. Suppose that the real values of non-deferrable demands turn out to be 11 for all time-slots (note that this realization falls within the uncertainty set, and there is no intra-day prediction in this example). The offline optimal peak for

this example is simply  $48/48 + 11 = 12$ . If we apply the SH-MPC algorithm to this example, the resulting peak will be 15.5 (see Fig. 2), and thus the worst-case competitive ratio would be no smaller than  $15.5/12 = 1.29$ . In contrast, applying the computational approach in Section III to this uncertainty set, the optimal competitive ratio can be shown to be 1.06, which is much smaller than 1.29 in the case of SH-MPC. As shown in Fig. 2, the poor performance of the SH-MPC algorithm can be understood as follows. In the first time-slot, SH-MPC sees that the current non-deferrable demand (i.e., 11) is higher than the predicted value (i.e., 10). However, SH-MPC still assumes that the future non-deferrable demand is the same as the predicted value (i.e., 10). Because the EV demand is flexible, SH-MPC computes a schedule that smoothens out the future demand. Thus, the charging rate at the time-slot 1 only increases slightly and the corresponding amount of EV demand is deferred to the future. This procedure is repeated, until towards the end of the horizon, SH-MPC realizes that the earlier predictions have been consistently wrong. However, now the decisions in the earlier time-slots cannot be reverted. SH-MPC has no choice but to increase the peak to accommodate the EV demand deferred from the past. In summary, since SH-MPC fails to account for the possible future deviations from the predicted values, it leads to poor performance in the worst case. In contrast, the methodology that we propose below will explicitly account for future uncertainty, and hence will provide much better worst-case guarantees.

### D. Summary of Notations

For ease of reference, we list all the notations in this paper in Table I.

## III. FUNDAMENTAL LIMIT OF THE COMPETITIVE RATIO

In this section, we extend the computation framework in [19] to find a fundamental lower bound on the competitive ratio  $\eta_Y(\pi)$  of any algorithm  $\pi$ . This lower bound will be given by the solution of the optimization problem (9). However, solving (9) is much more difficult than that in [19]. In Section III-B, we will develop a general convexification technique to convexify (9). Such a convexification technique generalizes fractional-linear programs [25], and thus may be of independent interest.

We need the following three lemmas throughout this section.

**Lemma 1:** Given a demand realization  $Z$ , a sufficient and necessary condition for a service profile  $E = [E_1, E_2, \dots, E_T]$  to be feasible, i.e., all demand can be completed before the corresponding deadlines, is that for all  $t_1 \leq t_2, t_1, t_2 \in \mathbb{T}$ , the following inequality holds,

$$\sum_{t=t_1}^{t_2} \sum_{s=t}^{t_2} a_{t,s} + \sum_{t=t_1}^{t_2} b_t \leq \sum_{t=t_1}^{t_2} E_t. \quad (5)$$

*Proof:* See Appendix A.  $\blacksquare$

Lemma 1 is a generalization of Lemma 6 in [19]. It states that, in order for a service profile  $E = [E_1, E_2, \dots, E_T]$  to be feasible, the total energy procured from the grid plus the renewable energy supply in any time interval  $[t_1, t_2]$  must be no smaller than the total demand that must be served in

TABLE I. LIST OF NOTATIONS.

$T$	Total Number of time-slots.
$\mathbf{a} = [a_{i,j}]$	A $T \times T$ matrix representing EV demand.
$\mathbf{b} = [b_i]$	A $T \times 1$ matrix representing net non-deferrable demand.
$\mathbf{x}(t)$	Demand variable revealed at time $t$ , including $a_{t,t}, \dots, a_{t,T}$ and $b_t$ .
$\hat{\mathbf{x}}^L(0, t), \hat{\mathbf{x}}^U(0, t)$	Lower bound and upper bound of $\mathbf{x}(t)$ . Known at time 0 (day-ahead prediction).
$\hat{\mathbf{x}}^L(u_t, t), \hat{\mathbf{x}}^U(u_t, t)$	Lower bound and upper bound of $\mathbf{x}(t)$ . Known at time $u_t$ (intra-day prediction).
$W(u_t, t)$	Maximum gap between $\hat{\mathbf{x}}^L(u_t, t)$ and $\hat{\mathbf{x}}^U(u_t, t)$ . Known at time 0 (day-ahead prediction).
$Y$	All quantities known at time 0, including $\hat{\mathbf{x}}^L(0, t), \hat{\mathbf{x}}^U(0, t)$ and $W(u_t, t)$ for any $t$ .
$Z$	A vector containing all the uncertain quantities not known at time 0.
$Z_t$	A sub-vector of $Z$ that contains the quantities revealed at or before time $t$ , including $[\mathbf{x}(s), s \leq t]$ and $[\hat{\mathbf{x}}^L(u_s, s), \hat{\mathbf{x}}^U(u_s, s), u_s \leq t]$ .
$Z_{>t}$	A sub-vector of $Z$ that contains the quantities revealed after time $t$ , including $[\mathbf{x}(s), s > t]$ and $[\hat{\mathbf{x}}^L(u_s, s), \hat{\mathbf{x}}^U(u_s, s), u_s > t]$ .
$E_t(Z_t, \pi)$	Total amount of energy drawn from the grid under an causal online algorithm $\pi$ .
$E_{\text{off}}^*(Z)$	Offline optimal peak energy consumption (assuming $Z$ is known at time 0, which is non-causal).
$\eta_Y(\pi)$	Competitive ratio of the online algorithm $\pi$ .

the same interval. Further, the condition (5) is also sufficient. Specifically, if we use the Earliest-Deadline-First (EDF) policy to serve the demand, then all the demand can be finished before the corresponding deadlines.

Clearly, Lemma 1 also applies to an online algorithm  $\pi$ . Specifically, a feasible online algorithm  $\pi$  must be able to support all possible demand trace  $Z \in \mathcal{Z}_Y$ . Then, according to Lemma 1, the service profile  $[E_t(Z_t, \pi), t = 1, 2, \dots, T]$  must satisfy (5) for all possible demand trace  $Z \in \mathcal{Z}_Y$ .

When studying online algorithms, we usually compare the peak of an online algorithm with the peak of an offline optimal algorithm. The following lemma gives a close-form formula of the offline optimal peak.

**Lemma 2:** Given a realization of  $Z$ , for any  $t_1 \leq t_2, t_1, t_2 \in \mathbb{T}$ , define the intensity of an interval  $J = [t_1, t_2]$  as

$$g_J(Z) = \frac{\sum_{t=t_1}^{t_2} (\sum_{s=t}^{t_2} a_{t,s} + b_t)}{t_2 - t_1 + 1}. \quad (6)$$

Then, the offline optimal peak is given by

$$E_{\text{off}}^*(Z) = \max\{0, \max_J \{g_J(Z)\}\}. \quad (7)$$

Lemma 2 states that the offline optimal peak is equal to

the maximum intensity over all possible intervals. The reason that we have a “0” term in Eqn. (7) is that the power procured from the grid must be non-negative. Lemma 2 is easy to prove. Clearly, the offline optimal peak must be no smaller than the maximum intensity given by (7). Otherwise, the demand inside the interval with the maximum intensity cannot be finished before the end of the interval. Further, there exists an offline solution with a peak exactly equal to  $E_{\text{off}}^*(Z)$ . Specifically, we can construct an offline solution that always procure  $E_{\text{off}}^*(Z)$  amount of energy at each time-slot. According to Lemma 1, the service profile of this offline solution can indeed finish all the demand before the corresponding deadlines.

#### A. Lower Bound

Consider an online algorithm  $\pi$  with competitive ratio  $\eta_Y(\pi)$ . We first study the maximum value for  $E_t(Z_t, \pi)$  given a realization  $Z$ . Recall that the decision  $E_t(Z_t, \pi)$  should only depend on  $Z_t$ . Further, we note that there may exist different realizations  $Z = \Pi_t([Z_t, Z_{>t}])$  that yield the same value of  $Z_t$ , because the future  $Z_{>t}$  can vary given the past  $Z_t$  up to time  $t$ . Thus, the value of  $E_t(Z_t, \pi)$  must be chosen such that it is no greater than  $\eta_Y(\pi)$  times the offline-optimal peak for any possible future realization  $Z_{>t}$ . Let

$$E_t^{\text{pe}}(Z_t) = \inf_{Z' \in \mathcal{Z}_Y: Z'_t = Z_t} E_{\text{off}}^*(Z'), \quad (8)$$

where the superscript “pe” stands for “peak estimation”. Note that the infimum in (8) is taken over all possible realization  $Z' \in \mathcal{Z}_Y$  that has the same revealed components at time  $t$  as  $Z_t$ . Then, we have the following lemma. The detailed proof is in Appendix B.

**Lemma 3:** Given an online algorithm  $\pi$  with competitive ratio  $\eta_Y(\pi)$ , we must have  $E_t(Z_t, \pi) \leq \eta_Y(\pi) E_t^{\text{pe}}(Z_t)$ .

We now apply Lemma 1. If  $\pi$  is feasible, then for all  $Z \in \mathcal{Z}_Y$  and all  $t_1 \leq t_2, t_1, t_2 \in \mathbb{T}$ , we must have

$$\sum_{t=t_1}^{t_2} \left( \sum_{s=t}^{t_2} a_{t,s} + b_t \right) \leq \sum_{t=t_1}^{t_2} E_t(Z_t, \pi) \leq \eta_Y(\pi) \sum_{t=t_1}^{t_2} E_t^{\text{pe}}(Z_t).$$

Define the following optimization problem:

$$\eta_{t_1, t_2}^*(Y) = \sup_{Z \in \mathcal{Z}_Y} \frac{\sum_{t=t_1}^{t_2} \left( \sum_{s=t}^{t_2} a_{t,s} + b_t \right)}{\sum_{t=t_1}^{t_2} E_t^{\text{pe}}(Z_t)} \quad (9)$$

Let  $\eta_Y^* = \max_{t_1 \leq t_2, t_1, t_2 \in \mathbb{T}} \{\eta_{t_1, t_2}^*(Y)\}$ . Then,  $\eta_Y^*$  provides a lower bound on the competitive ratio, which is stated below.

**Theorem 4:** For any feasible online algorithm  $\pi$ , its competitive ratio must be no smaller than  $\eta_Y^*$ , i.e.,  $\eta_Y(\pi) \geq \eta_Y^*$ .

The above arguments share some similarity to Theorem 4 in [19]. However, computing  $\eta_Y^*$  here is much more difficult than that in [19]. The computation of  $\eta_Y^*$  requires solving the optimization problem (9). Like in [19], the denominator of the objective function in (9) is the optimal value of another optimization problem. In general, such a bi-level optimization problem is NP-hard [26]. In [19], special structures of the problem are exploited to convert a similar bi-level optimization problem to a convex problem, which is then easier to solve.

However, the techniques in [19] critically rely on the property that the uncertain quantities (i.e., the EV demand matrix) can be freely scaled up or down without violating system constraints. Unfortunately, this property does not hold in this paper. Specifically, if all the uncertain quantities in  $Z$  is component-wise multiplied by a large constant, it may violate the bounds from day-ahead prediction in (1). In the next subsection, we will develop a more general convexification technique than that in [19] to convexify the optimization problem (9).

### B. Convexification of Problem (9)

We present the key convexification technique in Lemma 5.

*Lemma 5:* Consider the following optimization problem:

$$\begin{aligned} M_1 = \sup_{\vec{x}, y} \quad & (c^T \vec{x} + \alpha)/y \\ \text{subject to} \quad & y = f(\vec{x}), A\vec{x} \leq b, \end{aligned} \quad (10)$$

where  $\vec{x}, c$  are  $n \times 1$  vectors,  $A$  is a  $m \times n$  matrix,  $b$  is a  $m \times 1$  vector, and  $\alpha, y$  are scalars. Suppose that the following two conditions hold:

- (a)  $f(\cdot)$  is a convex function of  $\vec{x}$ , and  $f(\vec{x}) > 0$  over the entire constrained region of  $A\vec{x} \leq b$ ;
- (b) There exists  $\vec{x}$  satisfying  $A\vec{x} \leq b$ , such that  $c^T \vec{x} + \alpha > 0$ .

Then, the optimal value  $M_1$  of (10) is equal to the optimal value  $M_2$  of the following optimization problem:

$$\begin{aligned} M_2 = \sup_{\vec{x}', u} \quad & c^T \vec{x}' + \alpha u \\ \text{subject to} \quad & 1 \geq u f(\vec{x}'/u), A\vec{x}' \leq bu, u > 0. \end{aligned} \quad (11)$$

*Remark 4:* The optimization problem (11) can be transformed from (10) as follows. First, we let  $\vec{x}' = \vec{x}/y, u = 1/y$ . Then, the resulting optimization problem will be similar to (11), except that we have a constraint  $1 = u f(\vec{x}'/u)$  instead of  $1 \geq u f(\vec{x}'/u)$ . Note that  $f(\vec{x})$  is a convex function.  $u f(\vec{x}'/u)$  must also be convex in  $(\vec{x}, u)$  because it is the perspective of  $f(\cdot)$  [27]. Therefore, after relaxing the constraint  $1 = u f(\vec{x}'/u)$ , the optimization problem (11) becomes a convex problem, and can be efficiently solved. The result of Lemma 5 can be viewed as a generalization of fractional-linear program [25], which requires  $f(\vec{x})$  to be linear. The detailed proof is available in Appendix C.

We are now ready to convexify (9). We assume that the condition (b) holds in our problem, i.e.,  $\sum_{t=t_1}^{t_2} (\sum_{s=t}^{t_2} a_{t,s} + b_t) > 0$  for some  $Z \in \mathcal{Z}_Y$ . In other words, we cannot serve all the EV demand and the background demand using only the renewable energy. This assumption is reasonable, because that the renewable energy is highly variable, and we need to procure energy from the external grid when the renewable energy turns out to be low. It remains to show that the condition (a) also holds for (9), i.e.,  $\sum_{t=t_1}^{t_2} E_t^{pe}(Z_t)$  is a convex function of  $Z$ . Obviously, it is sufficient to show that  $E_t^{pe}(Z_t)$  is a convex function of  $Z_t$ . The convexity of  $E_t^{pe}(Z_t)$  is ensured by the following lemma.

*Lemma 6:* Suppose that  $f(x, y)$  is a convex function defined on a convex set  $D$ . Let  $D_x = \{y : (x, y) \in D\}$ , then  $g(x) = \inf_{y \in D_x} f(x, y)$  is also a convex function.

Lemma 6 can be viewed as an extension of the result in Section 3.2.5 of [27, p. 87], with the difference that the set  $D_x$  now changes with  $x$ . Readers can refer to [28] for the complete proof of Lemma 6.

It only remains to prove the convexity of  $E_t^{pe}(Z_t)$ . Since  $Z = \Pi_t([Z_t, Z_{>t}])$ , we slightly abuse notation and write  $E_{\text{off}}^*(Z) = E_{\text{off}}^*(Z_t, Z_{>t})$ . We now view  $Z_t$  as  $x$ , and  $Z_{>t}$  as  $y$ . Then, based on (8), we can rewrite  $E_t^{pe}(Z_t)$  as  $E_t^{pe}(x) = \inf_{y: (x, y) \in \mathcal{Z}_Y} \{E_{\text{off}}^*(x, y)\}$ . The region of  $(x, y)$ , i.e.,  $\mathcal{Z}_Y$ , is a convex set because all the constraints in (1)-(3) are linear constraints. Further, according to (7), it is easy to verify that  $E_{\text{off}}^*(x, y)$  is a convex function of  $(x, y)$  over the convex set  $\mathcal{Z}_Y$ . Therefore,  $E_t^{pe}(Z_t)$  is a convex function.

*Remark 5:* Although the 2-IPM in this paper assumes only one intra-day prediction, the results in this section can also be generalized to general prediction models, with multiple intra-day predictions. Specifically, each intra-day prediction can be characterized by a similar set of inequalities as (2)-(3). As a result, the set  $\mathcal{Z}_Y$  of all possible realizations can still be described by a set of linear constraints. Then,  $E_t^{pe}(Z_t)$  is still a convex function based on Lemma 6, and thus the optimal competitive ratio in (9) can be computed efficiently using the convexification technique in Lemma 5. Further, due to the same reason, the Algorithm Robustification procedure to be developed in Section IV will also work for such more general prediction models.

## IV. ALGORITHM DESIGN AND ROBUSTIFICATION

Note that we have obtained a lower bound  $\eta_Y^*$  for the competitive ratio of any online algorithm, the next step is to design an online algorithm that can attain this lower bound. It turns out that we can use the idea of the EPS algorithm proposed in [19]. Specifically, at each time, an online algorithm can set  $E_t(Z_t, \pi) = \eta_Y^* E_t^{pe}(Z_t)$ . We also refer to this algorithm as the EPS (Estimated Peak Scaling) algorithm because it always scales up the estimated value  $E_t^{pe}(Z_t)$  of the lowest possible future peak by the competitive ratio  $\eta_Y^*$ . Like in [19], it is not difficult to prove from the definition of  $\eta_Y^*$  that this EPS algorithm is feasible for any input  $Z \in \mathcal{Z}_Y$  because the condition (5) is always satisfied. Thus, the EPS algorithm attains the optimal competitive ratio  $\eta_Y^*$ .

The problem of this EPS algorithm, however, is that although it achieves the optimal competitive ratio for the worst-case input, its average-case performance can be quite poor, i.e., its peak can be high for many other inputs. To understand this dilemma, note that according to Lemma 3, any online algorithm with optimal competitive ratio  $\eta_Y^*$  should set  $E_t(Z_t, \pi)$  to be no larger than  $\eta_Y^* E_t^{pe}(Z_t)$ . In the case of the EPS algorithm, it always set  $E_t(Z_t, \pi)$  to the highest possible value. Thus, it can be viewed as the most *conservative* algorithm. If the future input indeed followed the worst-case, such conservatism would have been essential to attain the optimal competitive ratio: by serving more demand up-front, the EPS algorithm avoids a potentially large peak in the future. However, if the future input is different from the worst case, the EPS algorithm will likely be *too conservative*. For example, if the future input followed precisely the one that produces the value  $E_t^{pe}(Z_t)$



in (8), then using a rate  $E_t(Z_t, \pi) = E_t^{pe}(Z_t)$  would have been sufficient. Thus, one could argue that, since the worst-case perhaps occurs very rarely, using EPS may turn to be a poor choice in most scenarios.

This conflict between worst-case performance and average-case performance is not uncommon in the context of competitive online algorithms [17]. An algorithm designed for the worst-case can exhibit poor average-case performance, making it less appealing for practical implementation. Ideally, we would like to design an algorithm with both good worst-case and good average-case performance. In the rest of this section, we will present a novel “robustification” procedure to design such an algorithm. Our key idea is as follows. We first identify not one, but a class of algorithms that all attain the optimal competitive ratio. Then, starting from any algorithm with reasonable average-case performance, we “robustify” its decision by comparing it to the above class of algorithms. The resulting algorithm will then achieve both the optimal competitive ratio and good average-case performance.

#### A. Online Algorithms with the Optimal Competitive Ratio

Suppose that  $\pi$  is an optimal online algorithm with competitive ratio  $\eta_Y^*$ . For any realization  $Z \in \mathcal{Z}_Y$ , we next study all possible values of  $E_t(Z_t, \pi)$  that the algorithm  $\pi$  can take. The upper bound on  $E_t(Z_t, \pi)$  is given by Lemma 3, i.e.,

$$E_t(Z_t, \pi) \leq \eta_Y^* E_t^{pe}(Z_t). \quad (12)$$

Next, we derive a lower bound for  $E_t(Z_t, \pi)$ .

At time  $t$ , we use  $r_{t,t_1}$  to represent the total not-yet-served demand with deadline no greater than  $t_1$ , which includes all the remaining demand (with deadline no greater than  $t_1$ ) from the previous time-slots, the newly arrived net demand, and the newly arrived EV demand with deadline no greater than  $t_1$ . Consider any time instant  $t_1 \geq t$ , given any input  $Z$  with the first part being  $Z_t$ , we must have

$$E_t(Z_t, \pi) + \eta_Y^* \sum_{s=t+1}^{t_1} E_s^{pe}(Z_s) \geq r_{t,t_1} + \sum_{s=t+1}^{t_1} \left( \sum_{w=s}^{t_1} a_{s,w} + b_s \right). \quad (13)$$

Here,  $a_{s,w}$  and  $b_s$  are the elements of  $Z$ . The right hand side is the total demand that has to be served within  $[t, t_1]$ , while the left hand side is the maximum possible energy procurement from the grid (assuming that each future energy procurement rate is set to the upper bound in (12)).

We move the term “ $\eta_Y^* \sum_{s=t+1}^{t_1} E_s^{pe}(Z_s)$ ” from the left-hand-side of (13) to the right-hand-side, and then we have

$$E_t(Z_t, \pi) \geq r_{t,t_1} + \sum_{s=t+1}^{t_1} \left( \sum_{w=s}^{t_1} a_{s,w} + b_s - \eta_Y^* E_s^{pe}(Z_s) \right). \quad (14)$$

Note that (14) must hold for all possible future inputs. Define the following optimization problem that maximizes the right hand side of (14) over all possible future inputs:

$$\sup_{Z' \in \mathcal{Z}_Y : Z'_t = Z_t} \sum_{s=t+1}^{t_1} \left( \sum_{w=s}^{t_1} a'_{s,w} + b'_s - \eta_Y^* E_s^{pe}(Z'_s) \right) \quad (15)$$

where  $a'_{s,w}, b'_s$  are the corresponding elements of  $Z'$ . Let  $R_{\eta_Y^*}^*(Z_t, t_1)$  be the optimal value of (15). Then, in order to attain the optimal competitive ratio, the following must hold

$$E_t(Z_t, \pi) \geq r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1). \quad (16)$$

Finally, the above inequality must hold for all  $t_1 \geq t$ . Therefore, we obtain the following lower bound for  $E_t(Z_t, \pi)$ :

$$E_t(Z_t, \pi) \geq \max_{t_1 \geq t} \{r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1)\}. \quad (17)$$

*Remark 6:* Note that  $E_s^{pe}(Z'_s)$  is a convex function (see Section III-B). Therefore, the objective of (15) is a concave function. Further, both constraints ( $Z' \in \mathcal{Z}_Y$  and  $Z'_t = Z_t$ ) of (15) are linear constraints. Hence, (15) is a convex optimization problem, and thus can be efficiently solved.

We summarize the above discussion into Lemma 7.

*Lemma 7:* For any feasible  $\eta_Y^*$ -competitive online algorithm, we must have

$$\max_{t_1 \geq t} \{r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1)\} \leq E_t(Z_t, \pi) \leq \eta_Y^* E_t^{pe}(Z_t).$$

*Remark 7:* We note a key difference in the qualitative nature of the upper and lower bounds. The upper bound of  $E_t(Z_t, \pi)$  depends only on the optimal competitive ratio  $\eta_Y^*$  and the past realization  $Z_t$ , but is independent of the past decisions  $E_s(Z_s, \pi), s < t$ . In contrast, the lower bound of  $E_t(Z_t, \pi)$  also depends on the past energy procurement  $E_s(Z_s, \pi), s < t$ . Due to this reason, the lower bound is more *adaptive*: if the energy procured from the grid in the previous time-slots is large, we will have less remaining demand  $r_{t,t_1}$ , and thus have a smaller value for the lower bound. Such an ability to adjust based on the past decisions is the key reason that we can robustify an algorithm with good average-case performance to have optimal competitive ratio.

**Input:** Time-slot  $t$ , the remaining demand  $r_{t,t_1}$  and the part  $Z_t$  that has been revealed.

- 1 Compute the lower bound (17) and upper bound (12), and let  $E_t(Z_t, \pi)$  be any value in between.
- 2 The aggregator purchases  $E_t(Z_t, \pi)$  amount of energy from the external power grid, and uses the renewable energy and the purchased energy  $E_t(Z_t, \pi)$  to serve the existing demand. The aggregator first serves the background demand  $b_t$ , and then serves the deferrable demand by the *earliest deadline first* (EDF) policy (i.e., demand with earlier deadline gets served first). The aggregator will stop serving demand if all the available demand at time  $t$  is completely served or the amount of energy  $E_t(Z_t, \pi)$  is exhausted.

#### Algorithm 1: A Class of Optimal Online Algorithms

Motivated by Lemma 7, we define a class of online algorithms, called ABS (Adaptive Bound-based Scheduling), in Algorithm 1. We first show that all ABS algorithms are well-defined. Specifically, we show that the lower bound (17) is always no greater than the upper bound (12). Therefore, it is always feasible to pick a value for  $E_t(Z_t, \pi)$  at each slot.



*Lemma 8:* Given  $Z \in \mathcal{Z}_Y$  and an algorithm  $\pi$  in the class ABS, at each time-slot  $t$ , we must have

$$\max_{t_1 \geq t} \{r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1)\} \leq \eta_Y^* E_t^{pe}(Z_t). \quad (18)$$

Lemma 8 is the key of this section, and its proof is non-trivial. We can see that both sides of (18) depend on  $\eta_Y^*$ . In fact,  $\eta_Y^*$  is the smallest value such that (18) always holds. For any  $\eta < \eta_Y^*$ , it is possible to construct a case  $Z' \in \mathcal{Z}_Y$  such that  $\max_{t_1 \geq t} \{r_{t,t_1} + E_\eta^*(Z_t, t_1)\} > \eta E_t^{pe}(Z_t)$  for some  $t$ .

*Proof:* Recall that  $\max_{t_1 \geq t} \{r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1)\}$  is the smallest value of  $E_t(Z_t, \pi)$  that satisfies (14) for all possible  $t_1$ 's and all possible future realizations. Therefore, in order to prove (18), it suffices to show that  $\eta_Y^* E_t^{pe}(Z_t)$  also satisfies (14), i.e.,

$$\eta_Y^* \sum_{s=t}^{t_1} E_s^{pe}(Z_s) \geq r_{t,t_1} + \sum_{s=t+1}^{t_1} \left( \sum_{w=s}^{t_1} a_{s,w} + b_s \right), \quad (19)$$

for all  $t_1 \geq t$ , and all possible  $a_{s,w}$  and  $b_s$ .

We prove by induction on  $t$ . When  $t = 1$ ,  $r_{1,t_1} = b_1 + \sum_{s=1}^{t_1} a_{1,s}$ . Therefore, the right hand side of (19) is  $\sum_{s=1}^{t_1} (\sum_{w=s}^{t_1} a_{s,w} + b_s)$ . Based on the definition of  $\eta_Y^*$  in (9), (19) holds trivially for all  $t_1 \geq t$  when  $t = 1$ .

Assume that (19) holds for a given  $t$  and all  $t_1 \geq t$ . We will show that (19) holds for  $t+1$  and all  $t_1 \geq t+1$ . Note that

$$r_{t+1,t_1} = (r_{t,t_1} - E_t(Z_t, \pi))^+ + b_{t+1} + \sum_{s=t+1}^{t_1} a_{t+1,s},$$

where  $(x)^+ = \max\{x, 0\}$ ,  $(r_{t,t_1} - E_t(Z_t, \pi))^+$  is the remaining demand with deadline no greater than  $t_1$  and  $b_{t+1} + \sum_{s=t+1}^{t_1} a_{t+1,s}$  is the new arrival demand with deadline no greater than  $t_1$ . If  $(r_{t,t_1} - E_t(Z_t, \pi))^+ = 0$ , then (19) holds trivially based on the definition of  $\eta_Y^*$ . In the rest of this proof, we only need to consider  $(r_{t,t_1} - E_t(Z_t, \pi))^+ > 0$ . In this case,

$$r_{t+1,t_1} = r_{t,t_1} - E_t(Z_t, \pi) + b_{t+1} + \sum_{s=t+1}^{t_1} a_{t+1,s}.$$

We prove by contradiction. Assume that there exist  $\tilde{t}_1, \tilde{a}_{s,w}, \tilde{b}_s$ , such that  $(r_{t,\tilde{t}_1} - E_t(\tilde{Z}_t, \pi))^+ > 0$  and

$$\eta_Y^* \sum_{s=t+1}^{\tilde{t}_1} E_s^{pe}(\tilde{Z}_s) < r_{t+1,\tilde{t}_1} + \sum_{s=t+2}^{\tilde{t}_1} \left( \sum_{w=s}^{\tilde{t}_1} \tilde{a}_{s,w} + \tilde{b}_s \right).$$

Then,

$$\begin{aligned} & 0 \\ & < r_{t+1,\tilde{t}_1} + \sum_{s=t+2}^{\tilde{t}_1} \left( \sum_{w=s}^{\tilde{t}_1} \tilde{a}_{s,w} + \tilde{b}_s \right) - \eta_Y^* \sum_{s=t+1}^{\tilde{t}_1} E_s^{pe}(\tilde{Z}_s) \\ & = r_{t,\tilde{t}_1} - E_t(\tilde{Z}_t, \pi) + \tilde{b}_{t+1} + \sum_{s=t+1}^{\tilde{t}_1} \tilde{a}_{t+1,s} \\ & \quad + \sum_{s=t+2}^{\tilde{t}_1} \left( \sum_{w=s}^{\tilde{t}_1} \tilde{a}_{s,w} + \tilde{b}_s \right) - \eta_Y^* \sum_{s=t+1}^{\tilde{t}_1} E_s^{pe}(\tilde{Z}_s) \\ & = \sum_{s=t+1}^{\tilde{t}_1} \left( \sum_{w=s}^{\tilde{t}_1} \tilde{a}_{s,w} + \tilde{b}_s \right) - \eta_Y^* \sum_{s=t+1}^{\tilde{t}_1} E_s^{pe}(\tilde{Z}_s) \\ & \quad + r_{t,\tilde{t}_1} - E_t(\tilde{Z}_t, \pi) \\ & \leq R_{\eta_Y^*}^*(\tilde{Z}_t, \tilde{t}_1) + r_{t,\tilde{t}_1} - E_t(\tilde{Z}_t, \pi). \end{aligned}$$

The last inequality holds based on the definition of the optimization problem (15). The above derivation implies that

$$E_t(\tilde{Z}_t, \pi) < R_{\eta_Y^*}^*(\tilde{Z}_t, \tilde{t}_1) + r_{t,\tilde{t}_1},$$

which contradicts to our choice of  $E_t(\tilde{Z}_t, \pi)$ .

Hence, (19) holds for  $t+1$  and all  $t_1 \geq t+1$ . By induction, (19) holds for all  $t$ 's and  $t_1 \geq t$ . Thus, Lemma 8 holds. ■

Next, we show that all ABS algorithms are indeed optimal.

*Lemma 9:* Any algorithm  $\pi$  in the class of ABS is feasible and achieves the optimal competitive ratio of  $\eta_Y^*$ .

*Proof:* The proof is straightforward. First, based on the choice of  $E_t(Z_t, \pi)$ , it is easy to see that the peak of the algorithm  $\pi$  never exceeds  $\eta_Y^*$  times the offline optimal peak. Thus, the algorithm  $\pi$  is  $\eta_Y^*$ -competitive. Second, let  $t_1 = t$  in (16). It is easy to check that  $R_{\eta_Y^*}^*(Z_t, t) = 0$ . Therefore,  $E_t(Z_t, \pi) \geq r_{t,t}$ , which implies that no demand will violate its deadline at time  $t$ . This completes the proof. ■

## B. Algorithm Robustification

We have characterized the structure of optimal online algorithms. It only remains to find an online algorithm in ABS that also has good average performance. Our strategy is to take any algorithm with reasonable average-case performance, and convert it into one in the class ABS. We call this procedure *Algorithm-Robustification*. The Algorithm-Robustification procedure is formally stated in Algorithm 2. Specifically, Step 3 of the procedure states that, if  $E_t(Z_t, \pi)$  is between the upper bound and the lower bound, then we use the decision of the original algorithm  $\pi$ . Otherwise, we “robustify” the decision by setting  $E_t(Z_t, \pi_{\text{Robust}})$  to one of the bounds, so that the resulting “robustified” algorithm belongs to ABS. Intuitively, this procedure implies that for most inputs the robust version of  $\pi$  will likely behave in the same way as the original algorithm. Hence, the average-case performance will likely be similar. However, if there is a danger that the competitive ratio may be violated in the future, the robustified algorithm will then take the more conservative decision represented by the bounds.

**Input:** A realization  $Z \in \mathcal{Z}_Y$ , the optimal competitive ratio  $\eta_Y^*$  and any online algorithm  $\pi$ .  
**Output:** An optimal online algorithm  $\pi_{\text{robust}}$  and its schedules  $E_t(Z_t, \pi_{\text{robust}})$ .

```

1 for  $t = 1 : T$  do
2   Compute  $\alpha = \max_{t_1 \geq t} \{r_{t,t_1} + R_{\eta_Y^*}^*(Z_t, t_1)\}$ ,
    $\beta = \eta_Y^* E_t^{pe}(Z_t)$ , and the schedule  $E_t(Z_t, \pi)$  of the
   online algorithm  $\pi$ .
3   Set  $E_t(Z_t, \pi_{\text{robust}}) = M_\alpha^\beta(E_t(Z_t, \pi))$ , where
    $M_\alpha^\beta(x) = \max\{\min\{x, \beta\}, \alpha\}$ .
4 end

```

**Algorithm 2:** Algorithm-Robustification Procedure

In practice, in Section V-C, we will robustify a well-known online algorithm, called Shrinking Horizon Model Predictive Control (SH-MPC). The SH-MPC algorithm usually exhibits good average-case performance [24]. However, its worst-case competitive ratio can be very poor (see Section II-C). We then apply this Algorithm-Robustification procedure to the SH-MPC algorithm. This robustified SH-MPC algorithm will then achieve optimal competitive ratio in the worst case. Further, our numerical results demonstrate that the robustified SH-MPC algorithm achieves almost the same average-case performance as the SH-MPC algorithm.

### C. Accommodating Incorrect Predictions

Our prediction model in Section II-A has implicitly assumed that the day-ahead and intra-day predictions are always “correct.” This assumption implies that the upper and lower bounds of intra-day prediction are always within the upper and lower bounds of day-ahead prediction, and the real-time values are always within the upper and lower bounds or intra-day predictions. Then, under this assumption, Lemma 8 guarantees that the robustification procedure in Algorithm 2 will always work, i.e., the lower limit  $\alpha$  computed by Step 2 of Algorithm 2) is always no greater than the upper limit  $\beta$ .

What if the predictions are incorrect? In reality, the predicted bounds are based on some confidence intervals. Thus, there will always be some small probability that the future realization falls outside of these bounds. In that case, Lemma 8 may not hold, and we may have  $\alpha > \beta$  in Algorithm 2. Obviously, our robustification procedure will fail.

Interestingly, Lemma 8 also suggests a way to “fix” the robustification procedure when the above situation happens. Essentially, when  $\alpha > \beta$  in Algorithm 2, it implies that the originally-computed competitive ratio  $\eta_Y^*$  (assuming that the day-ahead prediction is correct) is no longer the correct competitive ratio for the amount of uncertainty faced by the aggregator. The value of  $\eta_Y^*$  will have to be increased. We note the monotonicity of  $\alpha$  and  $\beta$  with respect to  $\eta^*$ : it is easy to check that  $\alpha$  is a monotone decreasing function of  $\eta_Y^*$ , while  $\beta$  is a monotone increasing function of  $\eta_Y^*$ . Thus, if we keep increasing  $\eta_Y^*$ , eventually we can make  $\alpha \leq \beta$ . Based on this discussion, we add the following *parameter-tuning step* between Step 2 and Step 3 in Algorithm 2:

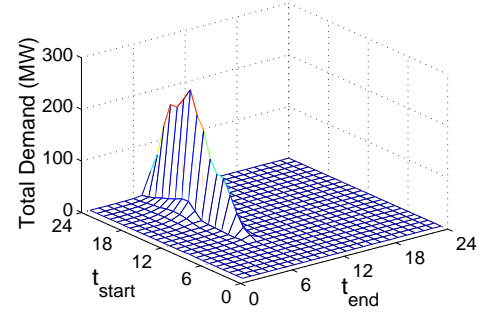


Fig. 3. Synthesized EV demand [29]. For any pair of discretized arrival time  $t_{\text{start}}$  and discretized departure time  $t_{\text{end}}$ , this figure plots the total EV demand.

**(Parameter-Tuning Step):** If  $\alpha > \beta$ , increase  $\eta_Y^*$  until  $\alpha \leq \beta$ .

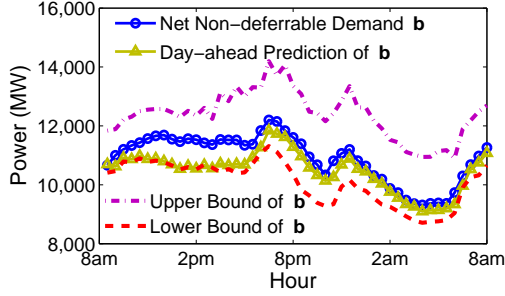
With the above parameter-tuning step, the algorithm robustification procedure can proceed even when the predictions are incorrect. Of course, incorrect predictions could negatively impact the performance of the system. We will study this impact using simulation in Section V-D.

## V. SIMULATION

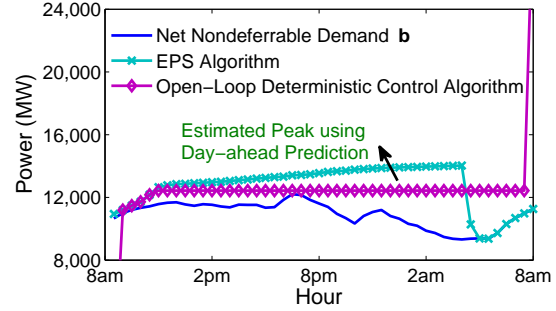
We conduct simulation using real traces from two data sets. Elia [30], Belgium’s electricity transmission system operator, provides day-ahead predictions and real-time values of background demand and renewable energy for every hour of each day. (However, Elia [30] does not provide data for intra-day prediction.) The National Household Travel Survey (NHTS) dataset [29] provides vehicle driving records for 150147 households. By assuming that future EV driving patterns are similar, it is not difficult to use the data in [29] to synthesize a model for the EV demand (see Fig. 3, and refer to our technical report [28] for more details), including EV arrival time, deadline and amount of energy charging demand, as has been done in earlier works in [31].

### A. The importance of Accounting for Uncertainty

We note that the day-ahead prediction in our 2-IPM consists of an upper bound and a lower bound for each time-slot. In contrast, the day-ahead prediction in Elia data-set [30] only contains one predicted value. Nonetheless, by comparing the difference between day-ahead predicted value and the real-time value over long periods of time (e.g., a year), it is easy to compute upper and lower bounds of the prediction error (for a given confidence level). Combining them with the day-ahead predicted values of [30], we can then generate the upper and lower bounds for day-ahead predictions as required in our model (see our technical report [28] for more details). In Fig. 4 (a), we apply this methodology to Elia’s data on background demand and renewable energy over a 24-hour period from 8am 02/05/2013 to 8am 02/06/2013, and plot the following versions of net non-deferrable demand **b** (as the background demand minus the renewable energy): the real-time value, the day-ahead predicted value directly from [30], and the upper and lower bounds of the real-time values as constructed above.



(a) Net non-deferrable demand and its day-ahead prediction (including both point prediction and interval prediction) from 8am 02/05/2013 to 8am 02/06/2013.



(b) The decision of the open-loop deterministic control algorithm vs. the EPS algorithm. The open-loop control algorithm produces a large peak, when the demand turns out to be higher in real time. In contrast, the peak of the EPS algorithm is much lower.

Fig. 4. The EPS algorithm vs. the open-loop deterministic control algorithm.

From Fig. 4 (a), we can see that the gap between the upper and lower bounds can be quite large (up to 20% of the day-ahead predicted value). The dataset in [30] does not provide explicit intra-day prediction. Hence, in our first experiment, we only consider day-ahead prediction. Lastly, for EV demand, we scale up<sup>1</sup> the synthesized model (see Fig. 3) by a factor 20, and assume that the day-ahead prediction of the EV demand is always accurate. In other words, we consider the uncertainty of background demand and renewable energy only.

We next demonstrate that, even for the scenario with low uncertainty, an algorithm that is oblivious to future uncertainty may lead to large peak consumption levels. Specifically, we consider the following open-loop deterministic control algorithm. At day-ahead, this algorithm assumes that the day-ahead predicted values of the background demand, the renewable energy (both from [30]) and the EV demand, are all accurate. It thus computes the offline optimal peak and the corresponding charging schedule (e.g., one possible schedule is to procure at each time-slot the amount of energy equal to this offline optimal peak), and then applies this schedule during real-time operation. Note that there is a chance that this schedule may not meet the deadline constraints of some EV demands because the real-time values will differ from the predicted values. In that case, this open-loop deterministic control algorithm will then need to procure additional energy at the time of the deadlines to meet the requirement of these EV demands. Intuitively, this open-loop deterministic control algorithm will perform poorly even if there is only a slight deviation between the real-time values and the predicted values because it always wait until the last minute to remediate the prediction error. This is confirmed from Fig. 4 (b), where we plot the energy procurement schedule of this open-loop deterministic control algorithm versus the EPS algorithm (discussed at the beginning of Section IV). The open-loop deterministic control algorithm suffers a large peak at the last minute because the deadlines

of most EV demands are 8am (see Fig. 3). In contrast, since the EPS algorithm increases the amount of energy procured early on, it avoids this last-minute peak. (We will see shortly that algorithms in the class of ABS will tend to have even lower peak than that of the EPS algorithm.) Hence, this figure clearly illustrates the importance of explicitly accounting for future uncertainty in the system.

#### B. 2-IPM and the Price of Uncertainty

We next evaluate the merit of the proposed 2-IPM in capturing the uncertainty of prediction. Note that given specific parameters of 2-IPM, we can calculate the lowest competitive ratio over all online algorithms (see Section III). This optimal competitive ratio can thus be viewed as a measure of the “price of uncertainty”, i.e., it represents the increase in cost (compared to the offline optimal peak) due to the inherent uncertainty captured by 2-IPM. Note that we have simulated based entirely on real traces in Section V-A. In the rest of the numerical experiments, we will manipulate the trace to observe the performance in different settings.

We first compare the competitive ratio under 2-IPM versus that under the prediction model in [19]. Note that the uncertainty model in [19] assumes that the ratio between the future uncertainty (i.e., the walk-in demand in [19]) and the predicted value (i.e., the reserved demand in [19]) is bounded. However, the absolute quantity of the predicted value is not specified. Thus, we refer to the uncertainty model in [19] as a *relative uncertainty model*. In contrast, in 2-IPM the absolute quantities for the predicted upper/lower bounds are specified. Hence, we refer to 2-IPM as an *absolute uncertainty model*. One can map absolute uncertainty in this paper to relative uncertainty in [19] by using only the ratio between the prediction error and the predicted value. For instance, suppose  $\mathbf{x}^{\text{DA}}(t)$  is the day-ahead predicted value. In 2-IPM, the upper and lower bounds of day-ahead prediction are specified as

$$\hat{\mathbf{x}}^L(0, t) = \mathbf{x}^{\text{DA}}(t) \times (1 - \epsilon), \hat{\mathbf{x}}^U(0, t) = \mathbf{x}^{\text{DA}}(t) \times (1 + \epsilon). \quad (20)$$

In contrast, with the relative uncertainty model in [19], only  $\epsilon$

<sup>1</sup>This EV trace [29] is obtained based on 150147 households. However, Belgium has 4 million households. Scaling the EV demand up by 20 will correspond to the future scenario where all vehicles in Belgium are electrified.

is specified, but not  $\mathbf{x}^{\text{DA}}(t)$ .

Intuitively, absolute uncertainty contains more information than relative uncertainty, and thus 2-IPM should yield lower competitive ratios. To confirm this point, we use the day-ahead predicted values as shown in Fig. 4 (a), but vary the upper/lower bounds of day-ahead prediction by varying  $\epsilon$  in (20). In Fig. 5, we plot the optimal competitive ratios under both 2-IPM and under the relative uncertainty model from [19], as  $\epsilon$  varies from 0.05 to 0.2. We can see that, even with only day-ahead prediction, the optimal competitive ratios under 2-IPM are lower. For example, when  $\epsilon = 0.2$ , the competitive ratio reduces from 1.2 to 1.16, which corresponds to approximately 4% reduction on the peak demand (which is significant as 1% reduction corresponds to  $0.01 \times 20\text{MW} \times \$9/\text{kW-month} \times 12 = \$21600$  saving per year for campus-level aggregators [8] with peak energy in the order of 20MW). In this sense, we argue that the price of uncertainty under 2-IPM is lower than that under a comparable model of relative uncertainty as in [19].

We next evaluate the impact of intra-day prediction. Note that the Elia data set [30] does not have intra-day prediction data. Thus, in the following we will artificially vary the parameters of intra-day prediction and evaluate the corresponding optimal competitive ratios. Such an evaluation methodology has a unique advantage: even before the operator carries out the intra-day prediction, our methodology will be able to reveal how useful such information will be in terms of reducing the optimal competitive ratio. Again, this knowledge of “price of uncertainty”, i.e., how much the cost can be reduced by intra-day prediction, could be very useful in deciding which types of intra-day prediction to perform and how accurate they need to be. Specifically, we evaluate three types of intra-day prediction, i.e., hour-ahead prediction, 12-hour-ahead prediction and 18-hour-ahead intra-day prediction. For each type of intra-day prediction, we vary the intra-day prediction gap as

$$W(u_t, t) = \min\{2\epsilon_{\text{intra}} \times \mathbf{x}^{\text{DA}}(t), \hat{\mathbf{x}}^U(0, t) - \hat{\mathbf{x}}^L(0, t)\},$$

where  $\hat{\mathbf{x}}^L(0, t), \hat{\mathbf{x}}^U(0, t)$  are the day-ahead predicted bounds specified in (20), and  $\epsilon_{\text{intra}}$  is the parameter we can vary. In Fig. 5, we plot the corresponding optimal competitive ratios under several choices of  $\epsilon_{\text{intra}}$ , as the  $\epsilon$  (i.e., error of day-ahead prediction) varies from 0.05 to 0.2. We can make a number of interesting observations. First, even if the hour-ahead prediction is perfect (i.e.,  $\epsilon_{\text{intra}} = 0$ ), the optimal competitive ratio barely changes from the case with only day-ahead prediction. Intuitively, this is because the hour-ahead prediction is too late: most of the decisions have already been made well before such hour-ahead prediction becomes available. In contrast, a perfect 12-hour-ahead prediction reduces the optimal competitive ratio by 2%. Interestingly, even an imperfect 18-hour-ahead prediction can be very helpful. For example, when  $\epsilon = 0.2$ , 18-hour-ahead prediction with  $\epsilon_{\text{intra}} = 0.08$  reduces the optimal competitive ratio from 1.16 (no intra-day prediction) to 1.13, which is comparable to the gain from a perfect 12-hour-ahead prediction. In practice, the earlier the intra-day prediction is performed, the less accurate it will likely be. Thus, the results in Fig. 5 will allow the

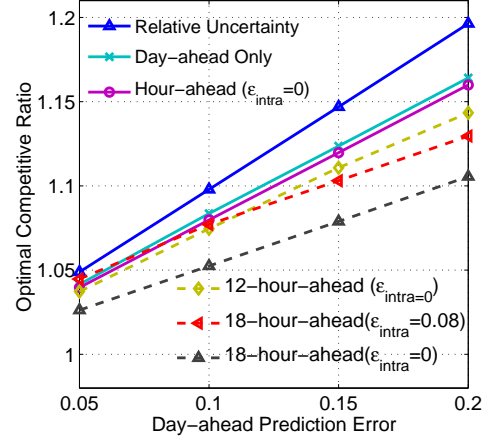


Fig. 5. Price of Uncertainty. Here, we compare the optimal competitive ratios between having only day-ahead prediction vs. having both day-ahead & intra-day predictions.

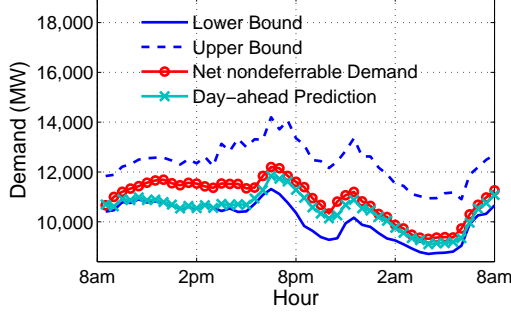
operator to evaluate which type of intra-day prediction will be most useful, i.e., in reducing the cost of uncertainty.

### C. Worst-case vs. Average-case Performance

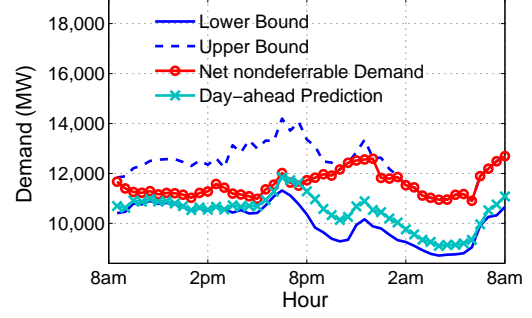
Until now we have focused on evaluating the worst-case competitive ratio. This worst-case competitive ratio is achievable by the EPS algorithm. However, as we discussed in Section IV, the EPS algorithm has poor average-case performance. In Section IV, we also present a robustification procedure that can be used to design algorithms with both good average-case performance and worst-case guarantees. Our next set of simulations will demonstrate this point.

Specifically, we will robustify a well-known heuristic algorithm, called Shrinking Horizon Model Predictive Control (SH-MPC) [24]. Empirically, the SH-MPC algorithm is often found to exhibit good average-case performance, especially when the future values of uncertain quantities are close to the predicted values. Nevertheless, we have also constructed a scenario in Section II-C where the SH-MPC algorithm performs much poorer than the optimal competitive ratio achieved by the EPS algorithm.

We next show that the robustified version of the SH-MPC algorithm (according to Section IV-B), will achieve both good worst-case and average-case performance. We will use two traces (see Fig. 6). In both traces, the day-ahead predicted values of background demand and renewable energy, and their corresponding upper-bounds and lower-bounds, are the same and are obtained using the methodology in Section V-A (see Fig. 4 (a)). Both traces also employ the same intra-day prediction model that uses the values of the respective quantities one time-slot ahead as the slot-ahead prediction for the next time-slot, and the intra-day prediction gap  $W(u_t, t)$  is set according to the maximum difference between the corresponding quantities in adjacent time-slots (see our technical report [28] for more details). Further, both traces use the same

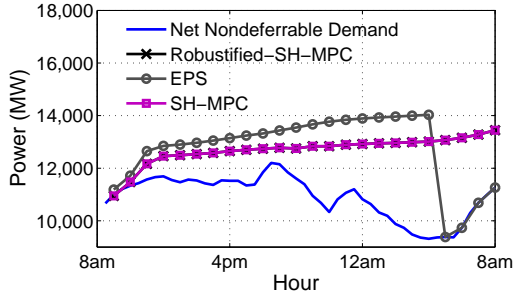


(a) Easy trace. The day-ahead predicted values are close to the real-time values of the net non-deferrable demand.

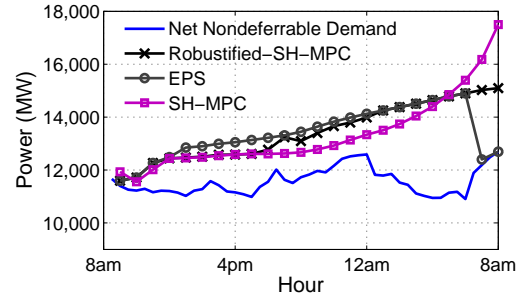


(b) Difficult trace. The day-ahead predicted values deviates significantly from the real-time values of the net non-deferrable demand.

Fig. 6. Net non-deferrable load of two Simulation traces.



(a) Easy trace. Note that the curve for “robustified-SH-MPC” overlaps with that of “SH-MPC”. Compared to the SH-MPC algorithm and the robustified-SH-MPC algorithm, the EPS algorithm produces an unnecessarily larger peak.



(b) Difficult trace. The SH-MPC algorithm cannot handle large uncertainty gracefully, and may lead to a large peak towards the end. Our robustified-SH-MPC algorithm can take actions earlier to prevent a large peak in the future.

Fig. 7. Schedules under two Simulation traces.

EV traces. Specifically, we use the EV demand shown in Fig. 3 as the day-ahead predicted value, and assume that the real demand vary uniformly randomly between 0.8 to 1.2 times the day-ahead predicted value. (We do not use intra-day prediction for EV demand.) However, the two figures differ in their revealed values of the net non-deferrable demand. In Fig. 6 (a), the revealed values of the net non-deferrable demand are closer to their day-ahead predicted values, while in Fig. 6 (b), the difference is much bigger (particularly at the end of the time-horizon). We will also refer to the trace in Fig. 6 (a) as the “easy trace”, and the trace in Fig. 6 (b) as the “difficult trace”.

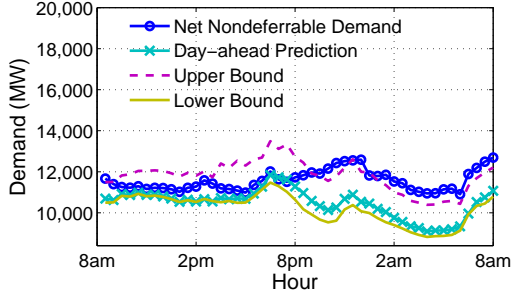
In Fig. 7, we compare the schedules of the EPS algorithm, the SH-MPC algorithm and the robustified-SH-MPC algorithm under both traces. By comparing Fig. 7 (a) and 7 (b), we observe that the EPS algorithm cannot distinguish between the easy trace and the difficult trace, and its peaks are similar high in both traces. In other words, the EPS algorithm is too conservative: it treats every trace as the worst trace, and scales up  $E_t^{pe}(Z_t)$  by the maximum value  $\eta_Y^*$ . In contrast, the SH-MPC algorithm produces a much lower peak in the

easy trace, when the day-ahead prediction is fairly accurate. However, its performance in the difficult trace is very poor. In the difficult trace, the day-ahead predicted values consistently underestimate the net non-deferrable load. As a result, the SH-MPC algorithm sets its service rate too low at the beginning, and has to use a much higher rate when all the EV demand approaches the deadlines. Our robustified-SH-MPC algorithm, on the other hand, inherits the benefits of both the EPS algorithm and the SH-MPC algorithm. For the easy trace, the robustified-SH-MPC algorithm gives virtually the same schedule as the SH-MPC algorithm. For the difficult trace, the robustified-SH-MPC algorithm detects that the service rate of the SH-MPC algorithm is too low at about 6pm. It then increases the service rate afterwards, and avoids the potential peak in the end. In summary, the robustified-SH-MPC algorithm achieves both good average-case and good worst-case performance.

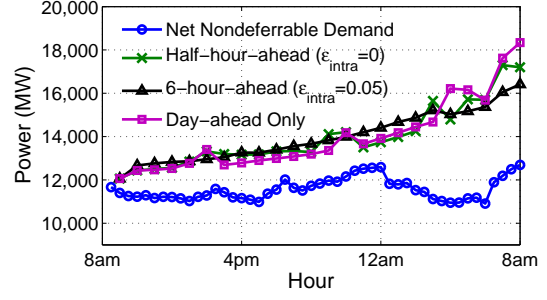
#### D. Impact of Incorrect Predictions

The above simulations have assumed that the predictions are always correct, i.e., future realizations always fall within the





(a) Difficult trace. (Note that the net-demand is higher than the day-ahead predicted upper bound after about 10pm.)



(b) Schedules of the Robustified-SH-MPC algorithm under 3 different parameter settings. Intra-day prediction becomes more useful when day-ahead prediction is wrong. For example, with 6-hour-ahead intra-day prediction, even though the optimal competitive ratio only reduces by 1.7% (from 1.081 to 1.063), the real peak is reduced by 10%.

Fig. 8. Impact of wrong day-ahead predictions.

bounds of earlier predictions, in which case the Robustified-SH-MPC algorithm achieves both efficiency and robustness. We next simulate a setting with incorrect day-ahead predictions, in which case we use the parameter-tuning step introduced in Section IV-C. Interestingly, our key observation below is that, although finer intra-day predictions play less of a role when all predictions are correct (see Fig. 5), they produce a bigger impact on the system performance when day-ahead predictions may be incorrect. Specifically, our simulation is based on the difficult trace in Fig. 6 (b), except that we tighten the day-ahead predicted upper/lower bounds to simulate an incorrect day-ahead prediction (see Fig. 8 (a)). Thus, the real net-demand may go beyond the bounds. The simulation setup for EV demand and intra-day prediction are the same as that in Section V-C and Section V-B, respectively. We choose different parameter settings for the intra-day prediction, and compare the performance of the Robustified-SH-MPC algorithm. From Fig. 8 (b), a key observation is that intra-day predictions reduce the peak significantly when day-ahead predictions are wrong. As a reference point for comparison, we compute the optimal competitive ratios under the three scenarios (day-ahead only, half-hour-ahead intra-day prediction with  $\epsilon_{\text{intra}} = 0$ , and 6-hour-ahead intra-day prediction with  $\epsilon_{\text{intra}} = 0.05$ ), and find them to be 1.081, 1.077, and 1.063, respectively. In other words, intra-day predictions only reduce the competitive ratio slightly when all day-ahead predictions are correct. In contrast, when day-ahead predictions are incorrect, the resulting peaks in Fig. 8 (b) can differ substantially. For example, if we have intra-day prediction that is 6 hours ahead with  $\epsilon_{\text{intra}} = 0.05$ , the peak can be reduced by as much as 10% compared to the scenario with only day-ahead prediction. This simulation result demonstrate that, even though intra-day predictions are not very effective in reducing the optimal competitive ratio, they are still very important to the overall robustness of the system, especially when day-ahead predictions are incorrect.

## VI. CONCLUSION

We study competitive online EV-charging algorithms for an aggregator to reduce the peak procurement from the grid. We model the uncertainty of the system using the 2-IPM, which captures both day-ahead and intra-day predictions of the demand and the renewable energy supply. We then develop a powerful computation approach that can compute the optimal competitive ratio under 2-IPM over any online algorithms, and also develop a class of online algorithms that can achieve the optimal competitive ratio. Noting that algorithms with the optimal competitive ratio (e.g., the EPS algorithm) may have poor average-case performance, we then propose a new *Algorithm Robustification* procedure that can convert an online algorithm with reasonable average-case performance to one with both the optimal competitive ratio and good average-case performance. We demonstrate the superior performance of such robustified algorithms via trace-based simulations.

There are a number of interesting directions for future work. First, we can study the impact of batteries on peak shedding. In this paper, we assume that EVs can only be charged. In contrast, dedicated batteries can not only be charged during idle hours, but also be discharged during peak hours. This additional flexibility could further reduce the peak demand. Second, in this work we have focused only on one aggregator. In practice, such an aggregator needs to participate in the overall electric power market. Then, it would be interesting to study the potential benefits of smart EV charging schemes for the entire power grid.

## ACKNOWLEDGEMENTS

This work has been partially supported by the NSF grants (CCF-1442726 and ECCS-1509536), the National Basic Research Program of China (Project No. 2013CB336700) and the University Grants Committee of the Hong Kong Special Administrative Region, China (Theme-based Research Scheme Project No. T23-407/13-N and Collaborative Research Fund No. C7036-15G).

## REFERENCES

- [1] S. Zhao, X. Lin, and M. Chen, "Peak-Minimizing Online EV Charging: Price of Uncertainty and Algorithm Robustification," in *IEEE INFOCOM*, Hong Kong, China, April 2015.
- [2] J. Cottrell, F. Fortier, and K. Schlegelmilch, "Fossil fuel to renewable energy," Feb. 2015.
- [3] Department of Communications, Energy & Natural Resources, "Ireland's Transition to a Low Carbon Energy Future," Dec. 2015.
- [4] T. Mai, D. Sandor, R. Wiser, and T. Schneider, "Renewable electricity futures study. executive summary," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep., 2012.
- [5] X. Hu, S. J. Moura, N. Murgovski, B. Egardt, and D. Cao, "Integrated Optimization of Battery Sizing, Charging, and Power Management in Plug-in Hybrid Electric Vehicles," *IEEE Transactions on Control Systems Technology*, vol. PP, no. 99, p. 1, 2015.
- [6] R. Sioshansi, "Modeling the impacts of electricity tariffs on plug-in hybrid electric vehicle charging, costs, and emissions," *Mathematics of Operations Research*, vol. 60, no. 2, pp. 1–11, 2012.
- [7] M. Liebreich, "Bloomberg new energy finance," *Bloomberg New Energy Finance Summit*, 2016.
- [8] <http://iitmicrogrid.net/>.
- [9] NationalGrid, "Understanding Electric Demand," [https://www9.nationalgridus.com/niagaramohawk/non\\_html/eff\\_elec-demand.pdf](https://www9.nationalgridus.com/niagaramohawk/non_html/eff_elec-demand.pdf).
- [10] Stem Inc., "Reducing demand charges," March 2015, [online] available at: <http://goo.gl/ofv414>.
- [11] F. Yao, A. Demers, and S. Shenker, "A Scheduling Model for Reduced CPU Energy," in *Proceedings of the IEEE symposium on Foundations of Computer Science*, Los Alamitos, CA, Oct. 1995.
- [12] L. Gan, U. Topcu, and S. H. Low, "Optimal Decentralized Protocol for Electric Vehicle Charging," *IEEE Transactions on Smart Grid*, vol. 28, no. 2, pp. 940–951, 2013.
- [13] B. Kouvaritakis and M. Cannon, *Model predictive control*. Springer, 2015.
- [14] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 1994.
- [15] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of Robust Optimization," *SIAM review*, vol. 53, no. 3, 2011.
- [16] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. on Power Systems*, vol. 28, no. 1, 2013.
- [17] S. Albers, "Competitive online algorithms," in *BRICS LS-96-2*, 1996.
- [18] N. Bansal, T. Kimbrel, and K. Pruhs, "Speed scaling to manage energy and temperature," *Journal of the ACM*, vol. 54, no. 1, March 2007.
- [19] S. Zhao, X. Lin, and M. Chen, "Peak-minimizing online EV charging," in *51st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, US, Oct. 2013.
- [20] B. Hodge and M. Milligan, "Wind Power Forecasting Error Distributions over Multiple Timescales," in *Power and Energy Society General Meeting*, Detroit, Michigan, July 2011.
- [21] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power System*, vol. 29, no. 3, pp. 1033–1044, 2014.
- [22] D. Saez, F. Avila, D. Olivares, C. Canizares, and L. Marin, "Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 548–556, 2015.
- [23] R. Hermans, M. Almassalkhi, and I. Hiskens, "Incentive-based coordinated charging control of plug-in electric vehicles at the distribution-transformer level," in *American Control Conference (ACC)*, 2012, pp. 264–269.
- [24] Z. K. Nagy and R. D. Braatz, "Nonlinear model predictive control for batch processes," *Control System Applications, Second Edition*, Edited by William S. Levine, CRC press, 2010.
- [25] A. Charnes and W. W. Cooper, "Programming with linear fractional functions," vol. 9, no. 3–4, pp. 181–186, 1962.
- [26] B. Colson, P. Marcotte, and G. Savard, "An Overview of Bilevel Optimization," *Annals of operations research*, vol. 153, no. 1, 2007.
- [27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2009.
- [28] S. Zhao, X. Lin, and M. Chen, "Robust Online Algorithms for Peak-Minimizing EV Charging under Multi-Stage Uncertainty," Technical Report. Available online: <https://engineering.purdue.edu/%7elinx/papers.html>.
- [29] U.S. Department of Transportation and Federal Highway Administration, "2009 National Household Travel Survey," <http://nhts.ornl.gov>.
- [30] <http://www.elia.be/en/grid-data/>.
- [31] D. Wu, D. C. Aliprantis, and K. Gkritza, "Electric energy and power consumption by light-duty plug-in electric vehicles," *IEEE transactions on power systems*, vol. 26, no. 2, pp. 738–746, 2011.

## APPENDIX

### A. Proof of Lemma 1

*Proof:* The necessity is obvious. We focus on the sufficiency in the following proof.

Suppose that the condition (5) holds for all  $t_1 \leq t_2, t_1, t_2 \in \mathcal{T}$ . Recall that  $Z$  contains entries that specify the EV demand, the net non-deferable demand, and the intra-day predictions. We will show that the earliest-deadline-first policy can finish all the demand specified in  $Z$  before the corresponding deadlines.

We prove by contradiction. Suppose that some demand misses its deadline. Without loss of generality, we assume that this demand's deadline is at time-slot  $d$ . We say a time-slot  $t < d$  is *good*, if and only if all the energy  $E_t$  is used to serve the demand with deadline no later than  $d$ . It is easy to see that the time-slot  $d$  is always good.

If all the time-slots  $t = 1, 2, \dots, d-1$  are good, then there is no energy wasted<sup>2</sup> during the first  $d$  time-slots, and all the energy is used to serve demand with deadlines no later than  $d$ . Note that  $\sum_{t=1}^d \sum_{s=t}^d a_{t,s} + \sum_{t=1}^d b_t \leq \sum_{t=1}^d E_t$ . Then, all demand with deadline no later than  $d$  can be finished before the end of the time-slot  $d$ , which contradicts to our assumption.

If there exists some time-slots  $t < d$  that is not good, let  $t_b = \max\{t < d | t \text{ is not good}\}$ . Then, in time-slots  $t = t_b + 1, \dots, d$ , no energy is wasted, and only demand with deadline smaller or equal to  $d$  is served. Furthermore, all demand with arrival time no later than  $t_b$  and deadline no later than  $d$  must have been completed before or at time-slot  $t_b$ . (Otherwise,  $t_b$  would have been good because the energy  $E_{t_b}$  could have been used to serve this part of demand according to the earliest-deadline-first policy.) Therefore, only demand with arrival time larger than  $t_b$  and deadline no later than  $d$  is served from  $t_b + 1$  to  $d$ . On the other hand, we note that  $\sum_{t=t_b+1}^d \sum_{s=t}^d a_{t,s} + \sum_{t=t_b+1}^d b_t \leq \sum_{t=t_b+1}^d E_t$ . Then, all demand with deadline no later than  $d$  can be finished before the end of the time-slot  $d$ , which contradicts to our assumption again. ■

<sup>2</sup>If at some  $t$ , the available demand to serve is less than  $E_t$ , we will say that part of  $E_t$  is wasted.



### B. Proof of Lemma 3

*Proof:* Recall the definition (8) that

$$E_t^{pe}(Z_t) = \inf_{Z' \in \mathcal{Z}_Y: Z'_t = Z_t} E_{\text{off}}^*(Z').$$

Since  $E_{\text{off}}^*(Z')$  is nonnegative,  $E_t^{pe}(Z_t)$  is also nonnegative, and thus is greater than  $-\infty$ . Then, for any  $\epsilon > 0$ , there must exist  $Z^\epsilon \in \mathcal{Z}_Y$  satisfying  $Z_t^\epsilon = Z_t$ , such that

$$E_{\text{off}}^*(Z^\epsilon) < E_t^{pe}(Z_t) + \epsilon.$$

Thus, in order for the online algorithm  $\pi$  to have a competitive ratio  $\eta_Y(\pi)$  for the input  $Z^\epsilon$ , its decision  $E_t(Z_t, \pi)$  must satisfy

$$E_t(Z_t, \pi) \leq \eta_Y(\pi) E_{\text{off}}^*(Z^\epsilon) < \eta_Y(\pi) (E_t^{pe}(Z_t) + \epsilon).$$

Letting  $\epsilon \rightarrow 0$ , we immediately have

$$E_t(Z_t, \pi) \leq \eta_Y(\pi) E_t^{pe}(Z_t).$$

■

### C. Proof of Lemma 5

*Proof:* We first show that  $M_2 \geq M_1$ . Let  $\{M_1^n\}$  be an increasing sequence satisfying  $M_1^n > 0$  for any  $n = 1, 2, \dots$ , and  $\lim_{n \rightarrow \infty} M_1^n = M_1$ . Since  $M_1 > 0$  (this can be easily proved based on the conditions (a) and (b) of the lemma), such a sequence  $\{M_1^n\}$  always exists. Recall that  $M_1$  is the optimal value of the optimization problem in (10). Then, for any  $n$ , there exists  $(\vec{x}_n, y_n)$  satisfying the constraints of (10), such that  $(c^T \vec{x}_n + \alpha)/y_n > M_1^n$ . Note that  $y_n = f(\vec{x}_n) > 0$  (by condition (a) of this lemma). Let

$$\vec{x}' = \frac{\vec{x}_n}{y_n}, u = \frac{1}{y_n}.$$

Then,  $(\vec{x}', u)$  satisfies the constraints in (11), and  $c^T \vec{x}' + \alpha u = (c^T \vec{x}_n + \alpha)/y_n > M_1^n$ . Noting that  $M_2$  is the optimal value of (11), we must have  $M_2 > M_1^n$ . Letting  $n \rightarrow \infty$ , we then have  $M_2 \geq M_1 > 0$ .

We next show that  $M_2 \leq M_1$ . Since  $M_2 > 0$ , there must exist an increasing sequence  $\{M_2^n\}$  satisfying  $M_2^n > 0$  for any  $n = 1, 2, \dots$ , and  $\lim_{n \rightarrow \infty} M_2^n = M_2$ . Then, according to the definition of  $M_2$ , for any  $n$ , there exists  $(\vec{x}'_n, u_n)$  satisfying the constraints of (11), such that  $c^T \vec{x}'_n + \alpha u_n > M_2^n > 0$ . Note that  $u_n > 0$ . Let

$$\vec{x} = \frac{\vec{x}'_n}{u_n}, y = \frac{1}{u_n}.$$

Then, it is easy to check that  $y \geq f(\vec{x})$  and  $A\vec{x} \leq b$ . Let  $y_0 = f(\vec{x}) \leq y$ . Then,

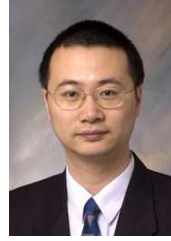
$$\frac{c^T \vec{x} + \alpha}{y_0} \geq \frac{c^T \vec{x} + \alpha}{y} = c^T \vec{x}'_n + \alpha u_n > M_2^n.$$

Noting that  $M_1$  is the optimal value of (10), we must have  $M_1 > M_2^n$ . Letting  $n \rightarrow \infty$ , we then have  $M_1 \geq M_2$ .

Combining the above analysis, we then have  $M_1 = M_2$ . ■



**Shizhen Zhao** received his B.S. from Shanghai Jiao Tong University, China in 2010, and Ph.D. degree from Purdue University, West Lafayette, IN, in 2015. He is currently working in Google Platform Networking team. His research interests are in the analysis, control and optimization in wireless networks, smart grid, and software defined networks.



**Xiaojun Lin** (S'02 M'05 SM'12) received his B.S. from Zhongshan University, Guangzhou, China, in 1994, and his M.S. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 2000 and 2005, respectively. He is currently an Associate Professor of Electrical and Computer Engineering at Purdue University.

Dr. Lin's research interests are in the analysis, control and optimization of wireless and wireline communication networks. He received the IEEE INFOCOM 2008 best paper and 2005 best paper of the year award from Journal of Communications and Networks. His paper was also one of two runner-up papers for the best-paper award at IEEE INFOCOM 2005. He received the NSF CAREER award in 2007. He was the Workshop co-chair for IEEE GLOBECOM 2007, the Panel co-chair for WICON 2008, the TPC co-chair for ACM MobiHoc 2009, and the Mini-Conference co-chair for IEEE INFOCOM 2012. He is currently serving as an Associate Editor for IEEE/ACM Transactions on Networking and an Area Editor for (Elsevier) Computer Networks Journal, and has served as a Guest Editor for (Elsevier) Ad Hoc Networks journal.



**Minghua Chen** (S'04 M'06 SM'13) received his B.Eng. and M.S. degrees from the Dept. of Electronic Engineering at Tsinghua University in 1999 and 2001, respectively. He received his Ph.D. degree from the Dept. of Electrical Engineering and Computer Sciences at University of California at Berkeley in 2006. He spent one year visiting Microsoft Research Redmond as a Postdoc Researcher. He joined the Dept. of Information Engineering, the Chinese University of Hong Kong in 2007, where he is currently an Associate Professor. He is also

an Adjunct Associate Professor in Institute of Interdisciplinary Information Sciences, Tsinghua University. He received the Eli Jury award from UC Berkeley in 2007 (presented to a graduate student or recent alumnus for outstanding achievement in the area of Systems, Communications, Control, or Signal Processing) and The Chinese University of Hong Kong Young Researcher Award in 2013. He also received several best paper awards, including the IEEE ICME Best Paper Award in 2009, the IEEE Transactions on Multimedia Prize Paper Award in 2009, and the ACM Multimedia Best Paper Award in 2012. He is currently an Associate Editor of the IEEE/ACM Transactions on Networking. He serves as TPC Co-Chair of ACM e-Energy 2016 and General Chair of ACM e-Energy 2017. His current research interests include energy systems (e.g., smart power grids and energy-efficient data centers), intelligent transportation system, online competitive optimization, distributed optimization, multimedia networking, wireless networking, and delay-constrained network coding.