2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Discount-based Pricing and Capacity Planning for EV Charging under Stochastic Demand

Parthe Pandit[1] and Samuel Coogan[2]

***Abstract*— The increasing market penetration of electric vehicles (EVs) poses new business avenues for existing facilities such as parking lots, gas stations, and other EV charging aggregators. There remain several open problems for these players in the EV power supply chain, such as pricing, scheduling the charging, and capacity planning, with limited theoretical understanding about their optimality. In this paper we consider an EV charging aggregator that provides energy to randomly arriving users with a user specified charging deadline, and we assume the aggregator has a total power budget which must be satisfied with high probability at each time instant.**

We look at a class of parameterized static pricing functions that incentivize EV users to provide the aggregators with longer deadlines, to avoid having to charge each EV at the peak kW rate. Under this pricing, we derive non-asymptotic concentration bounds on the required power capacity under stochastic arrival of users using a queuing theoretic approach.

***Index Terms*— Non-asymptotic concentration, Bernsteins inequality, $M/G/\infty$ queues, Static pricing.**

I. Introduction

The rapid influx of EVs in the automobile market poses several challenges for service providers in the energy supply chain. For example, power utility companies are concerned with planning generation-capacity for the grid to sustain a volatile electricity demand, minimizing overheads arising from this load variability, and pricing power for large scale users to correct the aforementioned problems without compromising on revenue.

On the other hand, public charging facilities such as parking lots at airports, shopping malls, restaurants or hotels, and DC fast charging facilities — hereafter together referred to as aggregators — sell energy to individual users and face decision making on two fronts, namely: (i) grid-to-aggregator, which includes planning power capacity for an anticipated demand, and (ii) aggregator-to-EV, which includes optimal pricing and efficient scheduling to prevent unreasonable service times. Several game theoretic and control theoretic formulations have been proposed previously to analyze and understand strategies that address these problems.

For example, as a model for individual users charging their EVs at night, [8] gives a non-cooperative game theoretic formulation for decentralized charging under shared power constraints. This work provides the existence of a Nash equilibrium charging strategy for each EV which nearly minimizes electricity generation costs by scheduling EV demand to fill the overnight non-EV demand *valley*. With a similar valley-filling objective, [3] considers an optimal-control framework and provides a decentralized algorithm for the overnight charging problem.

Similarly, [10] poses the grid-to-vehicle energy exchange as a noncooperative Stackleberg game in which the grid optimizes its revenue while the EVs choose their charging strategies resulting in a Stackleberg equilibrium. On the other hand, [4], [5] explore a pricing signal based distributed control for charging of EVs under shared constraints in the presence of an information-sharing network. On formulating each EVs decision-making process as a game, the authors give sufficient conditions on the pricing strategy of the aggregator under which this game has a unique Nash equilibrium. They also provide a distributed, consensus-based iterative algorithm that achieves this equilibrium.

The above works assume no temporal variability in the arrival and departure of EVs from the charging/discharging process, which prohibits their direct application to day-time public EV charging facilities such as parking lots of shopping malls, hotels, restaurants or airports, as well as DC fast-charging facilities. Under stochastic arrival of the users, [7] describes distributed scheduling policies with near-optimal aggregated performance that minimizes total cost of EV charging. We now briefly describe our main contributions.

A. Main Contributions

The current work proposes a problem formulation which aims to model a one-shot interaction between an aggregator and its EV users at the beginning of the energy transfer. Our formulation considers a stochastic arrival of EVs in the charging facility and assumes that the users arrive with predetermined energy demand. Using knowledge of the stochastic distribution of the energy demand, the aggregator broadcasts a static pricing function. Based on their impatience, each user chooses a service time deadline for the aggregator to fulfill its energy demand, such that the total cost, i.e., the monetary cost plus the opportunity cost of the service time for charging, is minimized.

In this setting, if the pricing function of the aggregator depends only on the energy demand, each user would choose the shortest feasible service time in order to minimize their opportunity cost. However in that case, the aggregator would be forced to have an instantaneous power capacity proportional to the product of the maximum power rate of

This work was supported in part by the UCLA *Sustainable LA Grand Challenge*.

[1]Parthe Pandit is with the Department of Electrical and Computer Engineering, at University of California, Los Angeles, parthepandit@ucla.edu

[2]Samuel Coogan is with the School of Electrical and Computer Engineering and the School of Civil and Environmental Engineering, at Georgia Institute of Technology, sam.coogan@gatech.edu

978-1-5386-5428-6/$31.00 ©2018 AACC

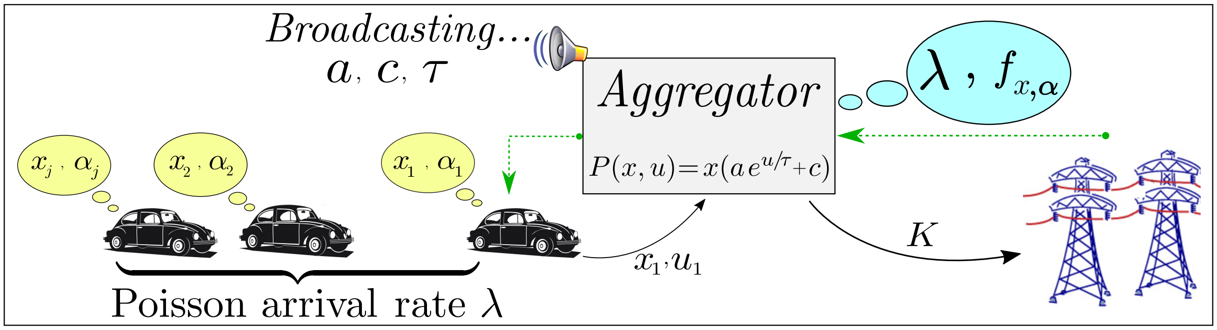


Fig. 1. Users arrive at the aggregator's facility with energy demand x_j and impatience factor α_j. Based on the information of the Poisson arrival rate λ, and statistics $f_{x,\alpha}$ of the users' demands and impatience factors, the aggregator chooses the parameters a, c, τ of the pricing function (2) and broadcasts them to the users. Each user submits its deadline u_j which minimizes its total cost, i.e., monetary cost plus opportunity cost given by equations (1) and (3). Subject to this optimal pricing, the aggregator also chooses a suitable power capacity K with the utility service so that constraint (PC) is rarely violated.

charging and the maximum number of active users expected for a given stochastic arrival rate.

Hence we look at a class of pricing functions which incorporate the service time deadline along with the energy demand, in order to incentivise users to choose suitably longer deadlines for the aggregator to fulfill the energy transfer. In general, the aggregator would allow impatient users to charge faster, but at a higher price, while offering discounted prices to patient users who are willing to provide longer deadlines. This aims to model the scenario where users could have varying impatience levels when charging at facilities near their work places, or in parking lots of shopping malls, or even gas stations equipped with charging capabilities.

With this discount-based pricing function, we prove that the requisite instantaneous power capacity for the aggregator is proportional to the sum of the maximum power rate of charging and the maximum number of active users.

More specifically, for a given stochastic arrival rate in EVs/hr., we provide high probability upper bounds on the number of active users and on the instantaneous power necessary to fulfill the demand. These upper bounds can aid the aggregators in making decisions regarding planning the spatial capacity and power capacity for a known arrival rate. We motivate the usefulness of high probability bounds for aggregators by explaining two possible scenarios — (i) In the *rare* event of the instantaneous power requirement exceeding our bound, the aggregator may contract some insurance with a third party that will provide the excess power to the aggregator. (ii) Alternatively, such probabilistic constraints could also help the aggregator plan for battery back-ups or gasoline powered generators in the *unfavourable* event of the power demand exceeding the capacity bound.

We simulate the performance of these bounds and observe that they are tight for a large range of arrival rates as well as confidence levels.

B. *Organisation*

We describe the exact model and the parametric pricing function in Section II. In Section III we formally present the main result of the paper in Theorem 3.1 and give its proof. Section IV demonstrates numerical results on simulated data which are coherent with Theorem 3.1. Section V discusses the generality of our problem formulation and pricing scheme to other resource aggregation problems catering to impatient users. Finally, we conclude the paper in Section VI.

II. Model description and Pricing

We consider an EV aggregator who has a facility to charge EVs. While charging the EVs, the aggregator obtains power from a utility service and supplies it to the EVs as depicted by the green dotten line in Figure 1.

A user j arriving at time a_j (in hr.), with demand x_j (kW-hr.), is allowed to choose a service time deadline u_j (hr.) for the aggregator to fulfill the energy transfer. In choosing the deadline, the user is faced with a pricing function $P(x_j, u_j)$ broadcast by the aggregator. Note that the users have stochastic but predetermined demands and arrival times, and only make decisions regarding the deadline u_j. We assume user j chooses u_j such that it minimizes the monetary cost $P(x_j, u_j)$ plus an opportunity cost $\alpha_j u_j$, i.e.,

$$u_j \in \underset{u \geq 0}{\operatorname{argmin}} \quad P(x_j, u) + \alpha_j u, \tag{1}$$

where α_j denotes the *impatience factor* (\$/hr.) of user j. We also assume that the demand x_j and α_j are bounded within the intervals $[x_{\min}, x_{\max}]$ and $[\alpha_{\min}, \alpha_{\max}]$ respectively, and these bounds are known to the aggregator.

Given the inputs x_j and u_j, the aggregator is tasked with charging EV j by x_j units of energy within the time interval $[a_j, a_j + u_j]$, in exchange for the payment $P(x_j, u_j)$. For simplicity, we assume that the users do not have to wait for an available charger. Furthermore, we assume that the aggregator employs a constant power charging strategy, i.e., the instantaneous power delivered by the aggregator to each user j at time t throughout its service time $t \in [a_j, a_j + u_j]$ is $\frac{x_j}{u_j}$ (in kW), whereby the energy transferred is x_j (kW-hr.).

We assume that from the aggregator's perspective, each user's predetermined preferences (x_i, α_i) are independent and identically distributed (i.i.d.) samples drawn from a distribution denoted $f_{x,\alpha}$, which is known to the aggregator, whereby u_j are also i.i.d.. We assume a Poisson arrival of EV users with rate λ (EVs/hr.), known to the aggregator.

In its transaction with the utility, the aggregator is faced with a constraint on the instantaneous power drawn, denoted $Q(t)$. We assume this constraint has a probabilistic form, i.e., for a confidence level $1 - \delta_K$ the aggregator is required to have, for all t,

$$\mathbb{P}\left(Q(t) := \sum_{i \in \mathcal{N}(t)} \frac{x_i}{u_i} \leq K\right) \geq 1 - \delta_K, \tag{PC}$$

where we define $\mathcal{N}(t) := \{i \mid t \in [a_i, a_i + u_i]\}$, the set of active users at time t.

A. Static Parametric Pricing

We consider a parametric pricing function P which aims to invite users to provide the aggregator with longer deadlines u_j. In particular, we study the following pricing function,

$$P(x, u) = x(a \cdot e^{-u/\tau} + c). \tag{2}$$

The incentive for the users is in the form of discounted rates per kW-hr for longer charging deadlines u. Parameter c is the *base rate*, a is the *surge price* (both in \$/kW-hr) and τ is the *effective service time* (in hrs).

Remarks: Observe that the function $P(x, u)$ is non-increasing and convex in the decision variable u, justified by a diminishing marginal discount for longer deadlines. Due to the convexity, (1) has a unique minimizer. Moreover using first order optimality conditions, for this parametric pricing function, the deadline and the payment can be expressed in closed form as

$$u_j = \tau \cdot \log\left(\frac{a x_j}{\alpha_j \tau}\right), \text{ and} \tag{3}$$

$$P(x_j, u_j) = c\, x_j + \alpha_j \tau, \tag{4}$$

where (4) is obtained on substituting (3) in (2). The peak rate a does not affect the revenue and base rate c does not affect the deadline decision. Both the deadline and payment are monotone increasing in a and c respectively. This may appear strange since the aggregator could potentially increase the parameters a and c indefinitely. However, practical considerations prevent the aggregator from doing so since for high prices, EV users will eventually balk and find a different aggregator.

In the analysis that follows, we assume the arrival rate λ of EV users is known to the aggregator, and c is fixed. We denote by $r_{\max}$ the maximum instantaneous rate (in kW) at which power can be transferred to an EV. This could indicate the power rating for a Level 3 DC Fast Charging unit.

In light of this maximum rate $r_{\max}$, the aggregator must choose (a, τ) such that $u_j \geq \frac{x_j}{r_{\max}}$ for all EVs j, in order to ensure each user provides the least feasible service time deadline.

Lemma 2.1: Suppose $\alpha_j \leq \alpha_{\max}$, and $x_{\min} \leq x_j \leq x_{\max}$. For any τ, if,

$$a \geq \alpha_{\max}\tau \cdot \max\left\{\frac{\exp\left(\frac{x_{\max}}{r_{\max}\tau}\right)}{x_{\max}}, \frac{\exp\left(\frac{x_{\min}}{r_{\max}\tau}\right)}{x_{\min}}\right\}, \tag{5}$$

then one can guarantee that $u_j \geq \frac{x_j}{r_{\max}}$ for all users j.

Proof: Observe that any $x_j \in [x_{\min}, x_{\max}]$ can be expressed as $x_j = \gamma x_{\min} + (1-\gamma) x_{\max}$, for some $\gamma \in [0, 1]$. Using Jensen's inequality for the $\log(\cdot)$ function, we have,

$$\begin{aligned} u_j := \tau \log \frac{a x_j}{\alpha_{\max}\tau} &\geq \tau\mu \log \frac{a x_{\min}}{\alpha_{\max}\tau} + \tau(1-\mu) \log \frac{a x_{\max}}{\alpha_{\max}\tau}, \\ &\geq \tau\mu \frac{x_{\min}}{r_{\max}\tau} + \tau(1-\mu)\frac{x_{\max}}{r_{\max}\tau} = \frac{x_j}{r_{\max}}. \end{aligned}$$

The last inequality follows from (5), proving the claim. ■

The above Lemma gives a method to choose a such that no user can choose a deadline infeasible for the fastest charging rate available. On the contrary if a user had choosen a deadline $u_j < \frac{x_j}{r_{\max}}$, then one can argue that there exists no scheduling strategy wherein the instantaneous power transferred to user j never exceeds $r_{\max}$.

III. Main Result

Our main result in this paper is the following theorem.

Theorem 3.1: Suppose users arrive at the aggregator at a Poisson arrival rate λ (EVs/hr.) and the maximum possible rate of charging is $r_{\max}$ (in kW). Assume that users have independent and identically distributed demands x_j and impatience factors α_j. Furthermore, each user chooses a deadline u_j given by equation (3) and is delivered power at a constant rate $\frac{x_j}{u_j}$ with expected value $\mathbb{E}\frac{x_j}{u_j} =: \mu$ and variance $\mathrm{Var}\left(\frac{x_j}{u_j}\right) =: \nu$. Then the following holds for a system at steady state.

(a) With confidence $1 - \delta_M$, the number of users in the system $N(t) := |\mathcal{N}(t)|$ will not exceed

$$M(\delta_M) = \lambda \cdot \mathbb{E}u + \frac{2}{3}\log\frac{1}{\delta_M} + \sqrt{\lambda \mathbb{E}u \log\frac{1}{\delta_M}},$$
$$\text{i.e.,} \qquad \mathbb{P}(N(t) \leq M(\delta_M)) \geq 1 - \delta_M. \tag{6}$$

(b) With confidence $1-\delta_K$, the instantaneous power drawn by the aggregator will not exceed

$$\begin{aligned} K(\delta_K) = \min_{\theta \in (0, \delta_K)} \Bigg\{ & M(\theta) \cdot \mu + \frac{2}{3} r_{\max} \log\frac{1}{\delta_K - \theta} \\ & + \sqrt{2\nu\lambda \mathbb{E}u \log\frac{1}{\delta_K - \theta}} \Bigg\}, \end{aligned}$$
$$\text{i.e.,} \qquad \mathbb{P}(Q(t) \leq K(\delta_K))) \geq 1 - \delta_K. \tag{7}$$

where $M(\theta)$ is as defined in part (a).

The above theorem prescribes two quantities which aid an aggregator plan its capacity both for physical space, as well as the instantaneous power capacity required to sustain the energy demand with the desired confidence. The *steady-state* assumption means that the arrival process has overcome the transients and the number of active users is well depicted by its stationary distribution.

The proof of Theorem 3.1 borrows ideas from non-asymptotic concentration inequalities in the context of queuing theory. As verified by our simulations presented in the later sections of the paper, the upper bounds for the confidence levels are reasonably tight and improve as λ increases.

Proof: To prove the theorem, we need to show (6) and (7).

(a) Observe that $N(t)$ is the number of users in the system of a steady state $M/G/\infty$ queue. Hence its stationary distribution is Poisson distributed with mean parameter $\lambda\mathbb{E}u$, (see [9]). Equation (6) follows by application of Corollary 1.2 detailed in Appendix.

(b) In order to prove (7), we first look at the following random variable, $Q(t)=\sum_{i\in\mathcal{N}(t)}\frac{x_i}{u_i}$, whereby on applying Bernsteins inequality from Lemma 1.1, we get,

$$\mathbb{P}\left[Q(t)>\mu N(t)+s|N(t)\right]\leq\exp\left(\frac{-\frac{1}{2}s^2}{N(t)\nu+\frac{sr_{\max}}{3}}\right),\quad(8)$$

where $\mu=\mathbb{E}\frac{x_j}{u_j}$. Observe that using the union bound gives,

$$\begin{aligned}\mathbb{P}\left(Q(t)>\mu\,M(\theta)+s\right)\leq\mathbb{P}\left(Q(t)>\mu\,N(t)+s\right)\\+\mathbb{P}(N(t)>M(\theta)).\qquad(9)\end{aligned}$$

By part (a), the second term in (9) is bounded by θ. In order to upper bound the first term in (9), we expand it using Bayes theorem as

$$\sum_{l=0}^{\infty}\mathbb{P}\left(Q(t)>N(t)\mu+s\mid N(t)=l\right)\mathbb{P}(N(t)=l),$$

which is less than $\sum_{l=0}^{\infty}\exp\left(\frac{-\frac{1}{2}s^2}{N(t)\nu+\frac{r_{\max}s}{3}}\right)\mathbb{P}(N(t)=l)$, using (8).

Observe that this is the expected value of the random function $\exp\left(\frac{-s^2}{2N(t)\nu+\frac{2r_{\max}s}{3}}\right)$, which by Jensen's inequality (see [2, Sec. 3.1.8]) is upper bounded by $\exp\left(\frac{-s^2}{2\mathbb{E}N(t)\cdot\nu+\frac{2r_{\max}s}{3}}\right)$. Consequently, we have,

$$\mathbb{P}\left(Q(t)>N(t)\mu+s\right)\leq\exp\left(\frac{-s^2}{2\mathbb{E}N(t)\cdot\nu+\frac{2r_{\max}s}{3}}\right).$$

Now, for $s^*:=\frac{2r_{\max}}{3}\log\frac{1}{\delta_K-\theta}+\sqrt{2\mathbb{E}N(t)\nu\log\frac{1}{\delta_K-\theta}}$, we get $\mathbb{P}(Q(t)>N(t)\mu+s^*)\leq\delta_K-\theta$. Together with (6) and (9) and using $\mathbb{E}\,N(t):=\lambda\mathbb{E}\,u$, gives

$$\mathbb{P}(Q(t)>K)=\mathbb{P}(Q(t)>M(\theta)\cdot\mu+s^*)\leq\delta_K.$$

Since this inequality holds for all $M(\theta)$ such that θ lies in the interval $(0,\delta_K)$, it also holds for $M(\theta^*)$ which gives the least upper bound K over this interval. Since the objective function blows up to $+\infty$ at $\theta=\delta_K$ and $\theta=0$, but is continuous and finite in the interval $\theta\in(0,\delta_K)$, it can be argued with ease that the infimum is achieved in this interval. This proves the claim. ∎

In the next section, we verify the performance of Theorem 3.1 on a simulated dataset. There, we observe that the performance of the bound K improves as the arrival rate λ increases.

IV. Simulations, Results and Discussions

A. Simulation Setup

In this section, we simulate an aggregator as shown in Figure 1 in order to verify the gap between the theoretical upper bounds from (6) and (7) predicted by Theorem 3.1 and the $1-\delta$ percentiles of the quantities $N(t)$ and $Q(t)$ at steady state. Let $\mathcal{N}^\delta[T_0]$ and $\mathcal{Q}^\delta[T_0]$ denote the $(1-\delta)^{\text{th}}$ percentiles of $N(t)$ and $Q(t)$ respectively, in a randomly chosen contiguous segment of duration T_0. More precisely, $\mathcal{Q}^\delta[T_0]$ is the *smallest* value for which

$$\frac{1}{T_0}\int_{t_0}^{t_0+T_0}\mathbb{I}_{\{Q(\tau)\geqslant\mathcal{Q}^\delta[T_0]\}}d\tau\leq\delta,$$

where $\mathbb{I}_{\{\cdot\}}$ is the indicator function. Note that t_0 is chosen uniformly at random. However, since we are interested in steady state behaviour, the choice of t_0 does not affect the statistics of these percentile quantities. $\mathcal{N}^\delta[T_0]$ is defined in a similar manner. We choose $T_0=8$ hr., to observe the behaviour in a typical work day, and for the rest of the paper, we drop the paranthesis $[T_0]$ while describing the quantities $\mathcal{N}^\delta$ and $\mathcal{Q}^\delta$. Recall that we need to verify that $M-\mathcal{N}^{\delta_M}$ and $K-\mathcal{Q}^{\delta_K}$ are almost always non-negative and that they are reasonably small with respect to M and K themselves.

To get a sufficient number of samples for $\mathcal{N}^\delta$ and $\mathcal{Q}^\delta$, we simulate 10 instances of the EV arrival process in 100 hr. long simulations. Since the arrival process is assumed to be Poisson with rate λ EVs/hr., for each of the 10 instances, N EV users arrive uniformly at random at the aggregator's facility during the interval $[0,100]$ hr, where $N\sim\text{Poisson}(100\cdot\lambda)$ independently for each instance. Each user exits the facility at the end of the service time deadline given by (3). We thus obtain

$$N(t)=\sum_{i=1}^{N}\mathbb{I}_{\{a_i\leq t\leq t+u_i\}},\text{ and }Q(t)=\sum_{i=1}^{N}\mathbb{I}_{\{a_i\leq t\leq t+u_i\}}\frac{x_i}{u_i}.$$

For the purpose of discretization during computation, these quantities are calculated at minute long intervals. From these quantities, we obtain 10 contiguous segments of duration 8 hr. to obtain 10 values of $\mathcal{N}^\delta$ and $\mathcal{Q}^\delta$ per instance. Note that the reason for choosing a 100 hr. long simulation for each of the 10 instances is not only to get sufficient number of samples for the behaviour in 8 hr. segments, but also to avoid the transient effects in the arrival process at the beginning of the instance. Since the transients are given by

$$\mathbb{E}[N(t)|N(0)=0]=\lambda\mathbb{E}u(1-e^{-t/\mathbb{E}u}),$$

(see [6, Sec. 5.2.2]) the transients exist for approximately $3\mathbb{E}u$. Therefore, we choose the starting point of each 8 hr. contiguous segment t_0 uniformly at random from the interval $[3\mathbb{E}u,(100-T_0)]$.

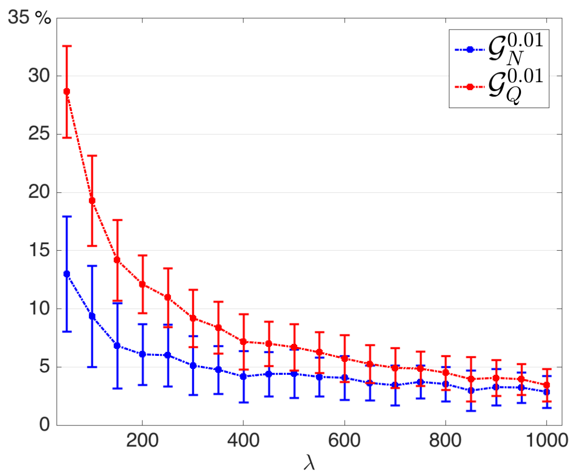


Fig. 2. The above figure presents the gaps measured using equation (10) for simulated quantiles $\mathcal{N}^{\delta_M}$ and $\mathcal{Q}^{\delta_K}$ with confidence parameters $\delta_M = \delta_K = 0.01$. The error bars indicate the standard deviation of the gaps $\mathcal{G}_N^{0.01}$ and $\mathcal{G}_Q^{0.01}$. Hence, for a confidence of 99% the upper bound tightens as the arrival rate increases.

B. Choice of parameters

We assume that $f_{x,\alpha}$ is a uniform distribution such that x_j is independent of α_j, where both x_j and α_j are uniformly distributed in the intervals $[x_{\min}, x_{\max}]$ and $[\alpha_{\min}, \alpha_{\max}]$ respectively. We simulate for the following choice of parameters: $x_{\max} = 100$, $x_{\min} = 10$ (in kW-hr.), and $\alpha_{\max} = 100$, $\alpha_{\min} = 10$ (in \$/hr). Since the choice of c does not affect the deadline decisions, we choose $c = 0$ for simplicity. However, in reality it would mimic the price for level 1 EV charging. Ideally, the aggregator can choose τ to meet the revenue targets, however we choose $\tau = 0.5$ hr., for which we get $\mathbb{E}u = 1.2$ hr. For the current choice of parameters $r_{\max}, x_{\min}, x_{\max}$ and $\alpha_{\max}$, the surge price $a = 5.3$\$/kW-hr. for this choice of τ, as indicated by Lemma 2.1.

For this choice of parameters, we get that u_j are distributed around the mean 1.21 hr. with a standard deviation 0.41 hr. Similarly, the power delivered to each user $\frac{x_j}{u_j}$ is distributed around its mean 45.07 kW with a standard deviation of 15.27 kW. Moreover, the maximum power rate delivered to a user is 122.75 kW.

C. Performance of bounds with varying arrival rate

The gap between the upper bounds given by Theorem 3.1 and the corresponding simulated quantiles $\mathcal{N}^{\delta_M}$ and $\mathcal{Q}^{\delta_K}$ is compared using the following metrics,

$$\mathcal{G}_N^{\delta_M} = \frac{M - \mathcal{N}^{\delta_M}}{M}, \text{ and } \mathcal{G}_Q^{\delta_K} = \frac{K - \mathcal{Q}^{\delta_K}}{K}, \tag{10}$$

with a lower value indicating better performance by the theorem. Figure 2 describes these metrics calculated for λ values ranging from 50 EVs/hr. to 1000 EVs/hr, $\delta_M = 0.05$, and $\delta_K = 0.05$. The monotonic decreasing nature of the curves indicates that $\mathcal{N}^{\delta_M}$ and $\mathcal{Q}^{\delta_K}$ are well estimated as the scale of the aggregators business increases.

While $\lambda = 1000$ EVs/hr. is a reasonable scale of arrival rates for parking lots of airports, etc., one might argue that the result in Theorem 3.1 seems too conservative for the scale of parking lots of shopping malls. However, consider a scenario where there are n aggregators facilities with each having an arrival rate of $\lambda_1, \lambda_2, \ldots, \lambda_n$. Recall that the net arrival rate of users at the n aggregators is given by $\lambda = \sum_{i=1}^n \lambda_i$, whereby, on sufficient scaling, aggregators can combine resources and use the results of Theorem 3.1 for planning their capacity.

D. Performance of bound with varying confidence levels

Figure 3 shows a comparison between M and $\mathcal{N}^{\delta_M}$, and K and $\mathcal{Q}^{\delta_K}$ for two values of λ. Observe that the bounds M and K are tight in the sense that an aggregator who plans for $\lambda = 150$ according to Theorem 3.1 would not be able to provide for an arrival rate of $\lambda = 200$ while maintaining the probabilistic constraint (7).

The space capacity requirement M prescribed by Theorem 6 is not overly conservative in the sense that if an aggregator plans according to $1 - \delta_M = 0.85$, they would not be able to satiate (6) with higher confidence $1 - \delta_M \geq .95$, as shown by the black dotted line in Figure 3. However, we see that the bound K varies significantly from $\mathcal{Q}^{\delta_K}$ in the high confidence regime $1 - \delta_K > .95$, i.e.Theorem 3.1 prescribes a capacity more conservative than may be necessary.

We now comment on the assumptions made in the analysis which may be relaxed and discuss the applicability of our framework to other similar resource aggregation problems catering to impatient users.

V. Remarks and Discussion

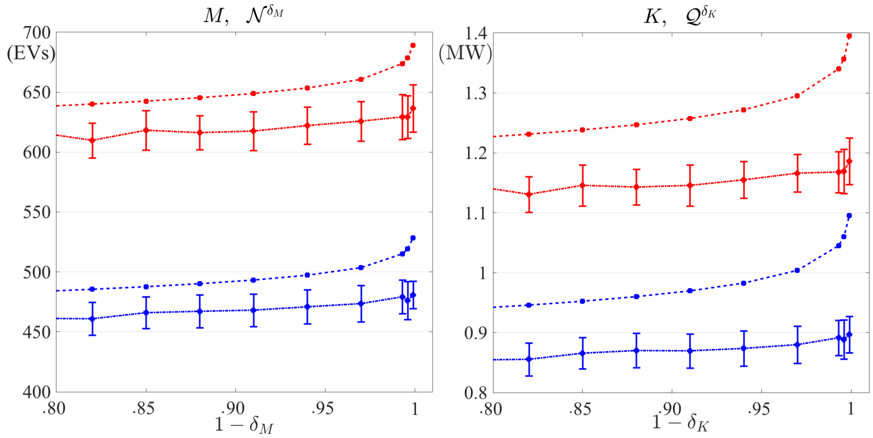
Through our simulations, we see that the first term dominates in the expression for the instantaneous power capacity K given in equation (7), i.e., for large enough λ, we have $K \propto \lambda \cdot \mathbb{E}u \cdot \mathbb{E}\frac{x}{u}$. Observe that this is lower bounded by $\lambda(\mathbb{E}\sqrt{x})^2$ using Holders inequality (see [2, Sec. 3.1.9]). Hence the optimal asymptotic rate is indeed linear in λ when using a constant power charging strategy.

In our analysis λ is assumed to be constant, while in reality the arrival rate could be modelled as a slowly varying non-homogenous Poisson process $\lambda(t)$. In such a situation, results from $M_t/G/\infty$ queues [9] could be invoked to obtain instantaneous power capacities $K(t)$.

For the high-confidence regime $1 - \delta > .97$, the bound on instantaneous power capacity given by Theorem 3.1 is conservative, suggesting that the rate of dependence of the prescribed capacity K with the confidence level δ_K could be suboptimal. We conjecture that this stems from the dependence of K on M, the maximum number of active users, instead of $\mathbb{E}Q(t)$ in (7).

Our problem formulation is fairly general and can be applied to other resource aggregation problems where an aggregator is distributing a *constrained* resource to impatient users looking to minimize their opportunity cost along with

Fig. 3. A comparison between the quantities M and $\mathcal{N}^{\delta_M}$, as well as K and $\mathcal{Q}^{\delta_K}$. While dotted lines correspond to the theoretical bounds M and K, the bold lines depict the mean of $\mathcal{N}^{\delta}_M$ and $\mathcal{Q}^{\delta_K}$, with the error bars highlighting their standard deviation. The blue plots correspond to $\lambda = 150$ EVs/hr., whereas the red plots correspond to $\lambda = 200$ EVs/hr.



the monetary cost. In such conditions, our pricing function is appropriate under suitable modifications. To enumerate a few examples of such aggregators with impatient users, consider cab aggregating services such as Uber and Lyft. In order to satiate the stochastic demand with fewer active cabs, these services provide users a cheaper alternative of sharing their rides (Uber Pool and Lyft Line respectively), in exchange for a slight delay in the transportation service. Similarly, services such as FedEx and Amazon offer discounted rates for delivery if the users choose longer service times. Another possible application of our model could be a discount-based pricing scheme for cloud-based computing services such as Amazon AWS, where users could choose a service time deadline if offered a cheaper rate per floating point operations (FLOPS).

VI. Conclusions

In this paper we presented a capacity planning problem faced by EV charging aggregators under stochastic energy demand and presented a static pricing function which incentives users to choose a longer service time deadline for the aggregator to perform the energy transfer. For the resulting deadline decisions by the users, with the help of nonasymptotic concentration inequalities, we give high probability upper bounds on the instantaneous power capacity required, as paraphrased in Theorem 3.1. This bound varies as $\mathcal{O}(r_{\max} + M)$, where $r_{\max}$ is the maximum allowable rate (in kW) for charging EVs and M is the maximum number of active users at the aggregators facility. In contrast, a service time invariant pricing scheme could have resulted in the aggregator requiring a power capacity proportional to $\mathcal{O}(r_{\max} \cdot M)$.

Numerical simulations show that this upper bound on the power capacity is tight with respect to the corresponding quantile of instantaneous power for a sizable range of the confidence levels. The power capacity prescribed by the bound gets tighter for larger arrival rates of users as shown in Figure 2.

Appendix

Lemma 1.1: (**Bernstein's inequality** [1, Cor 2.11]) Let X_i be independent random variables, with $|X_i - \mathbb{E}X_i| \leq b$, and $\sum_{i=1n}^{n} \mathrm{Var}(X_i) \leq n\vartheta$, then,

$$\mathbb{P}\Big(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq s\Big) \leq \exp\Big(\frac{-s^2}{2n\vartheta + 2bs/3}\Big).$$

Corollary 1.2: If X is a Poisson random variable with mean parameter λ, then for all $\epsilon > 0$, with probability at least $1-\epsilon$, we have $X \leq \lambda + \frac{2}{3}\log\frac{1}{\epsilon} + \sqrt{2\lambda \log\frac{1}{\epsilon}}$.

Proof: Observe that X is the limit of the sum $\sum_i X_i$ where $X_i \sim \mathrm{Ber}(p)$ such that $\lim_{n\to\infty} np = \lambda$. Drawing parallels with Lemma 1.1, b is 1 and $n\nu$ is $np(1-p) \leq np \to \lambda$. As result, $\mathbb{P}(X - \lambda > s) \leq \exp\Big(\frac{-s^2}{2\lambda + 2/3}\Big) \leq \epsilon$, where the last inequality is obtained by substituting $s = \frac{2}{3}\log\frac{1}{\epsilon} + \sqrt{2\lambda\log\frac{1}{\epsilon}}$, and using some elementary algebraic manipulations. ■

Lemma 1.3: ([9, Equation (23)]) The stationary distribution of the number of users in a $M/G/\infty$ queue, with poisson arrival λ and mean service time $\mathbb{E}S$ is poisson distributed with mean $\lambda \cdot \mathbb{E}S$.

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[2] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[3] L. Gan, U. Topcu, and S. H. Low. Optimal decentralized protocol for electric vehicle charging. *IEEE Transactions on Power Systems*, 28(2):940–951, 2013.

[4] B. Gharesifard, T. Başar, and A. D. Domínguez-García. Price-based distributed control for networked plug-in electric vehicles. In *American Control Conference (ACC), 2013*, pages 5086–5091. IEEE, 2013.

[5] B. Gharesifard, T. Başar, and A. D. Domínguez-García. Price-based coordinated aggregation of networked distributed energy resources. *IEEE Transactions on Automatic Control*, 61(10):2936–2946, 2016.

[6] D. Gross. *Fundamentals of queueing theory*. John Wiley & Sons, 2008.

[7] Y. He, B. Venkatesh, and L. Guan. Optimal scheduling for charging and discharging of electric vehicles. *IEEE transactions on smart grid*, 3(3):1095–1105, 2012.

[8] Z. Ma, D. S. Callaway, and I. A. Hiskens. Decentralized charging control of large populations of plug-in electric vehicles. *IEEE Transactions on Control Systems Technology*, 21(1):67–78, 2013.

[9] W. A. Massey. The analysis of queues with time-varying rates for telecommunication models. *Telecommunication Systems*, 21(2):173–204, 2002.

[10] W. Tushar, W. Saad, H. V. Poor, and D. B. Smith. Economics of electric vehicle charging: A game theoretic approach. *IEEE Transactions on Smart Grid*, 3(4):1767–1778, 2012.