Electrical Vehicle Charging Station Profit Maximization: Admission, Pricing, and Online Scheduling

Shuoyao Wang, Suzhi Bi, *Member, IEEE,* Ying Jun (Angela) Zhang, *Senior Member, IEEE,* and Jianwei Huang, *Fellow, IEEE*

***Abstract*—The rapid emergence of electric vehicles (EVs) demands an advanced infrastructure of publicly accessible charging stations that provide efficient charging services. In this paper, we propose a new charging station operation mechanism, the JoAP, which jointly optimizes the EV admission control, pricing, and charging scheduling to maximize the charging station's profit. More specifically, by introducing a tandem queueing network model, we analytically characterize the average charging station profit as a function of the admission control and pricing policies. Based on the analysis, we characterize the optimal JoAP algorithm. Through extensive simulations, we demonstrate that the proposed JoAP algorithm on average can achieve 330% and 531% higher profit than a widely adopted benchmark method under two representative waiting-time penalty rates.**

I. Introduction

ENVIRONMENTAL awareness and the rising fuel cost have stimulated an increasing interest in electrical vehicles (EVs). Establishing a conveniently available public charging infrastructure is essential to ensure a large market penetration of EVs [1]. Currently, however, many charging facilities are not yet profitable due to the low expected revenues, high capital expenditures, and high operating and maintenance costs [2].

In light of this, recent studies have focused on improving the operation efficiency of EV charging stations (e.g., [3]-[7]) by carefully designing the charging scheduling and pricing mechanisms. In particular, You and Yang in [3] characterized an optimal offline charging scheduling scheme, where "offline" means that the scheduling decision relies on the noncausal information of future EV charging profiles. Tang and Zhang in [4] relaxed the assumption of noncausal information by utilizing only the statistical distributions, instead of the exact realizations, of future EV charging profiles. In [5], Tang *et al.* designed an online charging scheduling algorithm that does not require any future information, not even the distributional information. Refs. [6] and [7] further proposed charging scheduling and pricing schemes to incentivize EV users to maximize the social welfare (i.e., minimizing the network-wide charging cost or maximizing the total economic surplus). Various pricing schemes have also been proposed to maximize the charging station's profit through time-scale decomposition, game theory, and dual decomposition [3]-[7].

Most existing studies, e.g, [3]-[6], assumed that a charging station has unlimited charging power to accommodate an infinite number of EVs simultaneously. In practice, however, the total charging power is bounded due to physical and security constraints of the distribution network. Moreover, the number of EVs that a charging station can accommodate is limited by the hardware and space constraints. As such, a charging waiting time (defined as the time between the arrival time of the EV to the charging station and the time that the EV starts to receive service) is unavoidable, which negatively impacts the users' experience. Hence, it is necessary to implement an effective admission control policy to reduce the impact of excessive charging waiting time due to random EV arrivals.

A practically-adopted admission control is the queue-length based admission (QBA) policy, where a newly arrived EV is admitted as long as the number of EVs waiting to be served at the station is below a specific threshold (e.g., the waiting room in the charging station). However, such a policy performs poorly in many cases, as illustrated in Section V. In contrast, Wei *et al.* in [8] proposed an admission control scheme, where the admission decision is based on the charging demands of EVs that have already arrived. The unknown future charging demands, however, were not considered in [8], resulting in poor profit performance in practical scenarios (see Section V for related examples).

In this paper, we propose a novel EV charging station operating mechanism that jointly optimizes pricing, charging scheduling, and admission control. The proposed algorithm, referred to as JoAP (joint admission control and pricing), maximizes the average profit of a charging station. Here, the profit is defined as the difference between the revenue and a penalty proportional to the average charging waiting time. The waiting time penalty reflects the EV owners' impatience of waiting in the queue for an excessively long time, which undermines the reputation of the charging station and reduces its long-term profit. In the JoAP algorithm, each EV user maximizes its surplus by adjusting its charging demand in response to the charging price. Meanwhile, the charging station maximizes its profit by choosing the proper admission control, scheduling and pricing policies.

The contributions of this paper are summarized as follows:

1) *Admission control, scheduling, and pricing scheme:* To the best of our knowledge, this is the first paper that

This work was supported in part by three grants from the Research Grants Council of Hong Kong under General Research Funding (Project number 2150828 and 2150876) and Theme-Based Research Scheme (Project number T23-407/13-N).

S. Wang, Y.J. Zhang, and J.W. Huang are with the Department of Information Engineering, The Chinese University of Hong Kong (E-mail: {ws013, yjzhang, jwhuang}@ie.cuhk.edu.hk@ie.cuhk.edu.hk). Y.J. Zhang is also with Shenzhen Research Institute, The Chinese University of Hong Kong.

S. Bi is with the College of Information Engineering, Shenzhen University, Shenzhen, Guangdong, China (E-mail: bsz@szu.edu.cn).

jointly optimizes pricing, scheduling, and admission control of an EV charging station. In particular, we propose a novel multi-sub-process based admission control scheme, which allows us to flexibly tradeoff between the revenue of the charging station and the waiting time of the EVs.

2) *Tandem queueing model:* We propose a tandem queueing model to analytically characterize the performance of the proposed JoAP algorithm. More specifically, we obtain closed-form expressions of the average waiting time and the admission probability as functions of the chosen algorithm parameters.
3) *Optimization of algorithm parameters:* Based on the tandem queue analysis, we propose a low-complexity algorithm to compute the close-to-optimal parameters of the JoAP algorithm. Our simulations show that JoAP algorithm on average achieves 330% and 531% higher profit than a widely adopted benchmark method under two representative waiting-time penalty rates. Moreover, by responding to the time of use (ToU) electricity price from the power grid, JoAP effectively reduces the peak load demand by 10.2%. Noticeably, a low peak load demand is essential to the safety, stability, and economic viability of power grids.

The rest of this paper is organized as follows. In Section II, we introduce the system model and formulate the problem. In Section III, we analyze the impact of the admission control policy on the admission probability and the average waiting time. In Section IV, we propose an efficient algorithm to simultaneously maximize the charging station's profit and individual EV user's payoff surplus. Simulation results are presented in Section V. Finally, we conclude this paper in Section VI.

TABLE I:
Key Notations

Sets	**Physical Meaning**
$\mathcal{V}$	The set of all EVs that arrive at the parking station during the time period of interest
Π	The feasible set of the admission control policies
Notations	**Physical Meaning**
α	Charging power
β	The elasticity parameter of the utility function $U(d)$
$\lambda_1, \lambda_2, \gamma$	The parameters of the mixture exponential distribution
$\omega_{\pi_n}(d)$	The average waiting time under π_n and d
π_n	A proposed admission control policy involving n sub-processes
ρ	Charging station load density
φ	EV Battery capacity
c	Waiting time penalty rate
d_i	The charging demand of EV i
$f_{\text{Ph}}(x)$	The PDF of the approximated inter-arrival time of admitted arrivals of Q.2
$F_X(x)$	The CDF of the inter-arrival time of admitted arrivals of Q.2
$f_X(x)$	The PDF of the inter-arrival time of admitted arrivals of Q.2
$h(\omega)$	The penalty when the average waiting time is ω
m	The number of charging ports
$P_j(n,d)$	The steady-state probability of state j under n and d
$P_{\pi_n}(d)$	The average admission probability under π_n and d
T_v	The minimum inter-arrival time of one sub-process
$U(d)$	The utility function of EVs
$x_i^{\mathcal{V}}(\pi_n, d)$	The binary admission decision of EV i
Variables	**Physical Meaning**
n	The number of sub-process in Q.1
r	The charging price announced by the charging station

II. System Model

A. Charging Station Operation

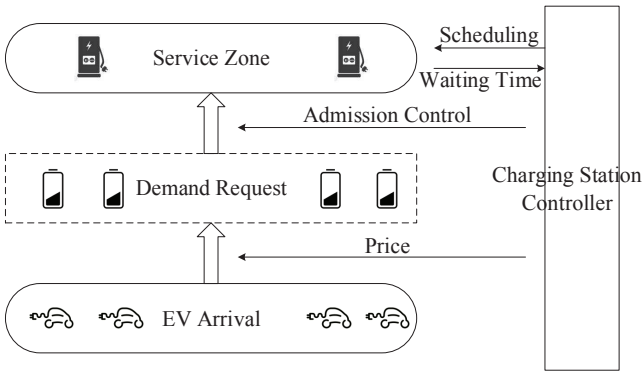


Fig. 1: The proposed charging station interaction system

We consider a charging station with m charging ports and a sufficiently large number of parking lots (i.e., much larger than m), as shown in Fig. 1. In this case, although a large number of EVs can be admitted to the charging station, at most m of them can be charged (served) simultaneously because of the physical constraints of the power distribution network and safety concerns. The charging ports are connected to the parking lots through a switch scheduler, which allows real-time communications and controls between a particular charging port and a scheduled EV. For the simplicity of analysis, we assume that the cost of connecting EVs with charging ports is negligible. All charging ports operate with the same fixed charging power α.

The charging station announces a charging price of r per unit energy to all arriving EVs. An EV i's payment to the charging station is the product of r and the EV's demand d_i. A long waiting time negatively affects the EV users' experience, which may lead to customer churn in the long run. Thus, the charging station aims to determine the optimal pricing and admission control policy to maximize its average profit, which is the revenue minus the penalty due to EVs' waiting.

EVs arrive at the charging station according to a Poisson random process [8][9], and each EV expects the charging station to fulfill its demand as soon as possible. When an EV i arrives, it attempts to maximize its surplus by choosing its charging demand d_i according to the charging price r. Based on the requested demand d_i, the charging station decides whether to admit the EV. The charging station optimizes the admission control policy to avoid excessive delay of admitted EVs. Once admitted, the EVs are charged on a first come first serve (FIFO) basis to fulfill their charging demands. It has been shown in [10] that, when all EVs have the same demand, the FIFO policy is equivalent to the shortest job first policy, and therefore is optimal in terms of minimizing the average

waiting time.

B. Optimization from EVs' Perspective

Suppose that each EV has a battery capacity φ and a utility function $U(d)$ that measures its satisfaction when a charging demand d is fulfilled. In general, $U(d)$ is an increasing concave function. Here, for simplicity, we adopt the following increasing concave function [11], where β is the elasticity parameter and $U(\varphi)$ is the maximum utility an EV can receive.

$$U(d) = U(\varphi)\frac{1-e^{-\beta d}}{1-e^{-\beta\varphi}}, \quad \forall 0 \le d \le \varphi. \tag{1}$$

Upon arrival, EV i determines its charging demand to maximize its customer surplus (i.e., utility minus payment) by solving

$$\max_{d_i} \quad U(d_i) - r d_i \tag{2a}$$

$$\text{s.t.} \quad 0 \le d_i \le \varphi. \tag{2b}$$

As Problem (2) is a concave maximization problem, we can compute the optimal demand d^* as a function of the service price r as follows.

$$d^*(r) = \begin{cases} -\frac{\ln(\frac{1-e^{-\beta\varphi}}{U(\varphi)\beta}r)}{\beta}, & \text{if } r \le \frac{U(\varphi)\beta}{1-e^{-\beta\varphi}}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

We can show that $d^*(r)$ is a decreasing function of the charging price r and becomes 0 when r is too high. This is consistent with the law of diminishing marginal utility (Gossen's First Law) in economics [12].

In this paper, we assume that the charging station knows the utility function (2). Accordingly, the station can predict EV's demand $d^*(r)$ in response to the price r as in (3). Thus, optimizing r is equivalent to optimizing d in the rest of the paper.

C. Optimization from Charging Station's Perspective

Let $\mathcal{V}$ denote the set of all EVs that arrive at the parking station during the time period of interest (e.g., 4 hours in our simulations). For each EV $i \in \mathcal{V}$, the charging station makes a binary admission decision $x_i^{\mathcal{V}}(\pi_n, d)$, where $x_i^{\mathcal{V}}(\pi_n, d) = 1$ if EV i is admitted, and $x_i^{\mathcal{V}}(\pi_n, d) = 0$ otherwise. Here, π_n denotes an admission policy, which will be detailed in Section III.A. Consequently, the average admission probability is

$$P_{\pi_n}(d) = \mathrm{E}_{\mathcal{V}}\left[\frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} x_i^{\mathcal{V}}(\pi_n, d)\right], \tag{4}$$

where $|\mathcal{V}|$ denotes the cardinality of $\mathcal{V}$. Moreover, the average waiting time achieved under policy π_n is a function of the demand d and the EV arrival process $\mathcal{V}$, denoted as $\omega_{\pi_n}(\mathcal{V}, d)$. Accordingly, the waiting time averaged over all the possible EV arrivals is denoted by

$$\omega_{\pi_n}(d) \triangleq E_{\mathcal{V}}[\omega_{\pi_n}(\mathcal{V}, d)]. \tag{5}$$

By satisfying an EV's charging demand d, the charging station receives a payment of rd, and pays an electricity cost of $p_e d$ to the utility company, where p_e is the ToU electricity price. The penalty related to the average waiting time is denoted by $h(\omega_{\pi_n}(d))$, where $h(\omega)$ is a general non-decreasing convex function of ω [13]. Based on this, we formulate the charging station's profit-maximization problem as follows.[1]

$$\max_{\pi_n, d} \quad P_{\pi_n}(d)\,(r - p_e)\,d - h\left(\omega_{\pi_n}(d)\right) \tag{6a}$$

$$\text{s.t.} \quad d \ge 0, \quad i \in \mathcal{V}, \tag{6b}$$

$$\pi_n \in \Pi, \tag{6c}$$

$$d = -\frac{1}{\beta}\ln\left(\frac{1-e^{-\beta\varphi}}{U(\varphi)\beta}r\right), \tag{6d}$$

where the feasible set Π will be introduced in Section III.A. The detailed expressions of $P_{\pi_n}(d)$ and $\omega_{\pi_n}(d)$ will be given in Section III.B and Section III.C, respectively.

III. Multi-Sub-Process Admission and Queueing Analysis

In this section, we first propose a multi-sub-process admission control scheme. Then, we present a tandem queueing model to analyze the impact of admission control policy and pricing decision on $P_{\pi_n}(d)$ and $\omega_{\pi_n}(d)$.

A. Admission Control and Queueing Model

The objective of admission control is to admit a large number of users with a guaranteed QoS. Let us first consider an extreme case of complete arrival-process regulation, i.e., the inter-arrival time of two successively admitted EVs is always larger than a predefined threshold as the result of the admission control. If such a threshold is large enough, then the waiting time of every admitted EV will be zero [10]. However, under this overly conservative admission control policy, the charging station utilization can be very low, hence not achieving the maximum profit. To achieve a good balance among the waiting time, admission probability, and server utilization, we propose a multi-sub-process admission control scheme consisting of n sub-processes. In particular, the inter-arrival time of two consecutively admitted EVs of the same sub-process must be larger than a threshold, denoted as T_v. An EV is admitted as long as it can fit in one of the sub-processes. With some abuse of notations, we use π_n to denote the proposed admission control policy involving n sub-processes. Hence, the feasible set of all admission control policies considered in this paper is $\Pi = \{\pi_n | n \in \mathcal{N}^+\}$.

Example 1. *Consider a π_2-admission policy that consists of two sub-processes, both having the same minimum inter-arrival time T_v, as shown in Fig. 2. When EV 1 arrives, we assign it to sub-process 1. When EV 2 arrives, we cannot assign to sub-process 1 as the inter-arrival time between EV 1 and EV 2 is shorter than T_v (the length of shadowed rectangle). Hence, we assign EV 2 to sub-process 2. For EV 3, we can assign it to sub-process 1. However, when EV 4 arrives, both sub-processes are "occupied". Therefore, EV 4*

[1] Problem (6) does not consider the penalty of denying the EVs. However, we can consider this by simply adding a linear term of $P_{\pi_n}(d)$ to the objective function. Doing so does not affect the structure of the problem, and our analysis will remain unchanged. For simplicity of exposition, this linear term is omitted for the time being.

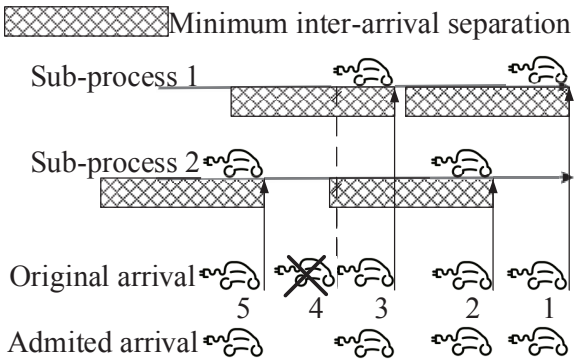


Fig. 2: Admission control example illustrated in Example 1

has to be rejected. When EV 5 *arrives, sub-process* 2 *becomes available again (due to the large enough inter-arrival time between EV 2 and EV 5). Hence, we accept EV* 5 *and assign it to sub-process 2.*

We would like to emphasize that the multi-sub-process scheme can represent a wide range of admission control policies. On one hand, the admitted traffic is completely regulated if there is only one sub-process, i.e., $n = 1$. On the other hand, when n approaches infinity, all EVs will be admitted regardless of the underlying distribution of the arrival process. Thus, choosing proper values of n and T_v allows us to balance the trade-off between the waiting time and admission probability, and eventually maximizes the charging station profit.

The admission process governed by the multi-sub-process scheme can be analyzed as a virtual queueing system with zero buffer and n servers, as represented by Q.1 (an $M/T_v/n/n$ queue[2]) in Fig. 3. Each virtual server corresponds to a sub-process, which has a deterministic service time T_v. The arrival of Q.1 is the EV arrival process $\mathcal{V}$. As the buffer is zero for Q.1, an EV will be declined for service if it finds all virtual servers are busy (i.e., all sub-processes are occupied) upon arrival. Otherwise, the EV is admitted and will occupy an idle virtual server for a fixed time period of T_v. The departure from Q.1 means that the EV is admitted to the charging station.

Once the EVs are admitted, they are served in the charging station according to the FIFO policy. We model the queueing system in the charging station as Q.2 in Fig. 3, where the m charging ports represent m servers, each with a deterministic service time d/α , where d and α are the charging demand per EV and the fixed common charging rate per charging port, respectively. Note that the departure process of Q.1 is the arrival process of Q.2. To ensure the stability of Q.2, the inter-departure time of Q.1 must be greater than the average service time of Q.2, i.e., $nT_v > md/\alpha$. We can equivalently represent this constraint as $nT_v = \tau md/\alpha$, where $\tau > 1$. To summarize, the determination of an admission control policy involves two decision variables, τ and n, with which the charging station can compute $T_v = \frac{\tau md}{n\alpha}$. In the following, we will consider optimizing n under a fixed value of τ. Without loss of generality, we assume that α equals 1. We will examine the impact of τ in Section V.

In practice, a well regulated arrival process seldom yields a long queue length [14]. Consequently, we ignore the impact of buffer of Q.2 and assume that it is infinite in the following analysis. In the remaining of this section, we are going to analyze the performance of the $\underbrace{M/T_v/n/n}_{\text{Q.1}} + \underbrace{\cdot/d/m}_{\text{Q.2}}$ tandem queueing network. The analysis will be useful in designing the JoAP algorithm in Section IV.

Before concluding this subsection, we would like to emphasize that Q.1 in Fig. 3 is a virtual queue that does not exist in reality. We consider Q.1 for the purpose of analyzing the admission control policy. Queue Q.2 is a real queue corresponding to the service in the charging station. As such, the admission probability is the probability that a new arrival is admitted to Q.1, and the charging waiting time is the waiting time in Q.2.

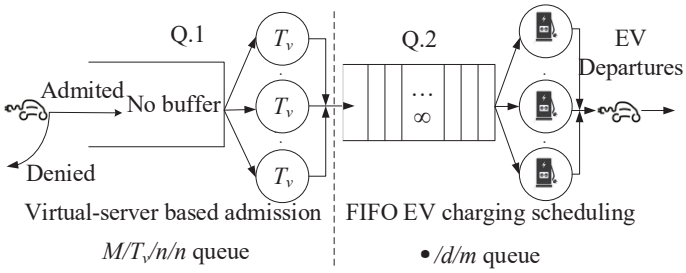


Fig. 3: The tandem queueing network model

B. Admission Probability

Previous queueing literature (e.g., [15]) has numerically analyzed the performance of $M/D/n/n{+}K$ queues (e.g., Q.1 in Fig. 3) without analytical characterization of the system performance. Tijms in [16] showed that a two-phase process server can be used to approximate a deterministic server with a marginal performance gap. Based on this approximate model, we derive a closed-form expression of steady-state probabilities of Q.1 in the following Lemma 1. To the best of our acknowledgment, this paper is the first analytical study of the $M/D/n/n{+}K$ system with $K = 0$ (i.e., zero buffer).

Lemma 1. *Consider an $M/D/n/n$ queue with a Poisson process with a arrival rate λ, a deterministic service time $D = \tau md/n$, and zero buffer-size. The steady-state probability of state j (i.e., the probability that the system has j users being served simultaneously) can be calculated based on the two-phase-process approximation in [16] as follows,*

$$P_j(n,d) = \frac{\left(\frac{d\tau m\lambda}{n}\right)^j}{j!\sum_{j'=0}^{n}\frac{\left(\frac{d\tau m\lambda}{n}\right)^{j'}}{j'!}}. \tag{7}$$

[2]We can represent a single queue using Kendall's notation in the form $A/S/C/C{+}K$, where A describes the inter-arrival times, S describes the service time, C describes the number of servers, and $C{+}K$ describes the number of spaces in the system. When the K parameter is not specified (e.g. $M/M/1$ queue), it is assumed that $K = \infty$. In Kendall's notation: M stands for Markov or memoryless process, D stands for deterministic process, G stands for general and corresponds to an arbitrary probability distribution, Ph stands for phase-type process (the process that constructed by a convolution or mixture of exponential process), and $\cdot$ stands for any process.

The admission probability of Q.1 is:

$$P_{\pi_n}(d) = 1 - P_n(n,d) = 1 - \frac{\left(\frac{d\tau m\lambda}{n}\right)^n e^{-\frac{\tau m d\lambda}{n}}}{\Gamma\left(n+1, \frac{\tau m d\lambda}{n}\right)}. \quad (8)$$

We prove Lemma 1 by induction. The detailed proof is given in Appendix A. The validity of Lemma 1 is verified in Fig. 4, where we compare the admission probability derived in (8)

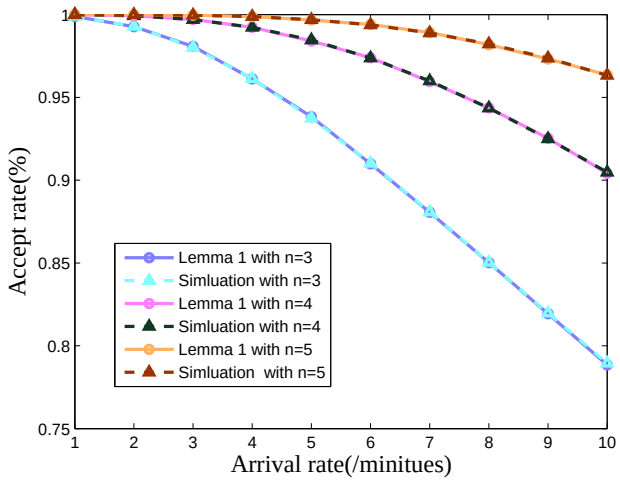


Fig. 4: The comparison of the admission probability between the simulation and the approximation in Lemma 1, with $\tau = 1.01$, $m = 4$, $\beta = 0.05$, $\alpha = 3.3kW$, $d = \varphi$, and $\gamma = 35kWh$

with the simulation results (without any approximation). We choose the number of servers in Q.1, n, to be 3, 4, and 5, respectively. Each point corresponds to the average over 1000 time periods. The maximum gap between the analysis and simulation is 0.01%, which verifies the accuracy of the results in Lemma 1.

C. Average Waiting Time

1) Admitted-arrival: To study the average waiting time in Q.2, we derive the PDF (probability density function) of the inter-arrival time of Q.2.

Lemma 2. *The PDF of the inter-arrival time of admitted arrivals of Q.2 is*

$$f_X(x) = \begin{cases} \sum_{i=0}^{n} \frac{i}{T_v} \left(\frac{T_v - x}{T_v}\right)^{i-1} P_i(n,d), & \text{if } x \le T_v, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Proof. Recall that the arrival process of Q.2 is the departure process of Q.1. According to [17], the residual service time of a queueing system is the service time remaining to a job under service when the system is observed at any time. The residual service time of Q.1 follows a uniform distribution in $[0, T_v]$, as the arrival process is memory-less (Poisson) and the buffer size is zero [18]. When Q.1 is at a particular state i, the probability of no departure during the next period of time of a length x is equal to the probability that the residual service times of all existing jobs are no-less than x, i.e., $\left(\left(T_v - x\right)/T_v\right)^i$. Consequently, the probability of the first departure time (after the observation time point) being no greater than x is $1 - \left(\left(T_v - x\right)/T_v\right)^i$. Therefore, the CDF of the inter-departure time of Q.1 (i.e., the inter-arrival time of Q.2), denoted by X, is,

$$F_X(x) = \begin{cases} \sum_{i=0}^{n} \left(1 - \left(\frac{T_v - x}{T_v}\right)^i\right) P_i(n,d), & \text{if } x \le T_v, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Taking the derivative of (10) yields the PDF in Lemma 2. □

2) Phase-type Approximation: We now derive the average waiting time of Q.2 with the phase-type approximation. So far, there does not exist a general closed-form expression for the waiting time distribution of a $GI/D/m$ queue (e.g., Q.2 in Fig. 3) [19], where GI means a general arrival process. To overcome this difficulty, [19] showed that the waiting time distribution of a $GI/D/m$ queue is the same as that of a $GI^{(m^*)}/D/1$ queue, where $GI^{(m^*)}$ denotes a coordinated inter-arrival time process that is distributed as the sum of m inter-arrival times of a $GI/D/m$ queue. Let Y denote the coordinated inter-arrival time of the $GI^{(m^*)}/D/1$. The mean and variance of X and Y are related by $\mu_Y = m\mu_X$ and $\sigma_Y^2 = m\mu_X^2 - m\mu_{X^2}$.

Furthermore, a $GI^{(m^*)}/D/1$ queue can be approximated by a $Ph/D/1$ queue, where Ph means the phase-type process [19]. One of the most widely used phase-type distribution is the mixture exponential distribution, which is defined as the mixture of two exponential distributions with means $1/\lambda_1$and $1/\lambda_2$, and weights γ and $1-\gamma$, respectively. Specifically, the PDF is given by

$$f_{\text{Ph}}(x) = \gamma e^{-\lambda_1 x} + (1-\gamma)e^{-\lambda_2 x}. \quad (11)$$

In this paper, we replace the inter-arrival distribution of Q.2 with the mixture exponential distribution in (11). To ensure that the first and second moments of the mixture exponential distribution are equal to those of Y, we set $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 2\mu_Y$, $\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} = \sigma_Y^2$, and $\gamma = \frac{1}{2}$. In this way, we can approximate the waiting time distribution of Q.2 by that of the $Ph/D/1$ queue.

We derive in the following Theorem 1 the approximated average waiting time of the charging station. In the theorem, we will use notation $\rho = \lambda P_{\pi_n}(d)d/m$ to denote the load density admitted to the charging station.

Theorem 1. *The approximated average waiting time at the charging station for the admitted EVs is*

$$\omega_{\pi_n}(d) = \frac{\rho d}{2(1-\rho)}\left[d^2 + 2d\mu_Y + \sigma_Y^2\right]. \quad (12)$$

Moreover, $\omega_{\pi_n}(d)$ is an increasing convex function in d for a fixed n.

Proof. We can derive the approximated average waiting time based on Proposition 1 quoted in [19].

Proposition 1. *[19] For a $Ph/D/1$ queue, let S and A denote the service time and the inter-arrival time, respectively. The Laplace transform $a^*(s) = \int_0^\infty e^{-st}a(t)dt$*

of the inter-arrival time A can be written as $a^*(s) = \frac{a_1(s)}{a_2(s)}$, where $a(t)$ denotes the probability density function of S and $a_1(s)$ and $a_2(s)$ are two polynomials. Then, the average waiting time can be approximated as $\frac{\rho E(S)}{2(1-\rho)}\left[E(S^2)+E(A^2)+2E(S)\frac{a_1^{'}(0)}{a_1(0)}-2\psi\frac{a_2^{'}(0)}{a_2(0)}\right]$, where $\psi = \frac{a_2^{'}(0)-a_1^{'}(0)}{a_2(0)}$ and ρ is the load density.

We apply Proposition 1 to our tandem queue model. As the Lapalaze transform of Y is $\mathcal{L}\{f_Y(x)\} = \frac{\lambda_1\lambda_2+\frac{1}{2}(\lambda_1+\lambda_2)}{(s+\lambda_1)(s+\lambda_2)}$, we have $a_1(s) = \lambda_1\lambda_2 + \frac{1}{2}(\lambda_1+\lambda_2)$, $a_2(s) = (s+\lambda_1)(s+\lambda_2)$, $\psi = \frac{\frac{1}{2}(\lambda_1+\lambda_2)}{\lambda_1\lambda_2}$. Taking the first order derivative of $\frac{\rho}{1-\rho}$ over d, we have $\frac{P^{'}d+P+\frac{\lambda}{m}P^2d}{(1-\rho)^2}$, which is a positive increasing function in d. Substitute λ_1, λ_2 with the representation of μ_Y and σ_Y, we can express the average waiting time as $d\left[d^2+EY^2+2d\frac{\frac{1}{2}\lambda_1+\lambda_2}{\lambda_1\lambda_2}-2\frac{\frac{1}{2}\lambda_1+\lambda_2}{\lambda_1\lambda 2}\frac{\lambda_1+\lambda 2}{\lambda_1\lambda 2}\right] = d\left[d^2+EY^2+2d\mu_Y-\mu_Y^2\right] = d\left[d^2+\sigma_Y^2+2d\mu_Y\right]$. Notice that $\mu_Y = \frac{m}{P_{\pi_n}(d)\lambda}$. Consequently, $d\left[d^2+\sigma_Y^2+2d\mu_Y\right] = d^3+2d\sigma_Y^2+\frac{md^2}{P_{\pi_n}(d)}$ is a convex increasing function in d for fixed n. Thus, $\omega_\pi(d)$ is a convex function in d, which is in agreement with Kingman's formula [20]. Kingman's formula is a well-known approximation for the mean waiting time in a $G/G/1$ queue, which states that $w_\pi(d) \approx \frac{\rho d}{2(1-\rho)}\left[d^2+EY^2\right]$. □

Let us verify the approximation by comparing the average waiting time in (12) with simulation results (without any approximation). In Fig. 5, for each pair of arrival rate and individual demand, we simulate 1000 independent 1000-hour arrival processes $\mathcal{V}$ and plot the average admission probabilitys. The difference is no more than 0.1%.

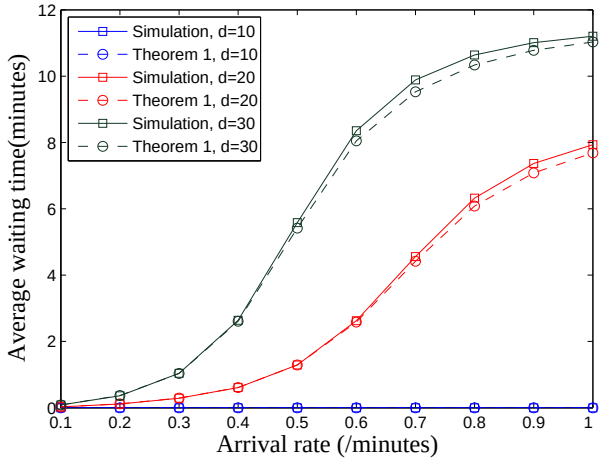


Fig. 5: The comparison of the average waiting time between the simulation and the approximation in Theorem 1, with $m = 4$.

IV. Optimization Problem Recasting and Profit Maximization

With the tandem queueing analysis, we rewrite (6) as

$$\max_{n,d} \quad s(n,d) = P_{\pi_n}(d)\left(\frac{de^{-\beta d}}{\xi} - dp_e\right) - h\left(\omega_{\pi_n}(d)\right) \tag{13a}$$

$$\text{s.t.} \quad 0 \le d \le \varphi, \quad n \in \mathcal{N}^+, \tag{13b}$$

where $\xi = \frac{1-e^{-\beta\varphi}}{U(\varphi)\beta}$, $P_{\pi_n}(d)$ and $\omega_{\pi_n}(d)$ are given in (8) and (12), respectively. Recall that n is the number of sub-processes in Q.1 and d is the estimated EV demand in response to the charging price r. Variables n and d together determine the admission control and pricing strategy of the charging station. To solve the integer programming Problem (13) efficiently, we replace the decision variable n with $P \triangleq P_{\pi_n}(d)$. This is because for a particular feasible (P,d), we can find a unique n that satisfies equation (8). Accordingly, (13) can be equivalently expressed as

$$\max_{P,d} \quad \hat{s}(P,d) = P\left(\frac{de^{-\beta d}}{\xi} - dp_e\right) - h\left(\omega_{\pi_n}(d)\right) \tag{14a}$$

$$\text{s.t.} \quad 0 \le d \le \varphi, \quad P \in (0,1), \tag{14b}$$

$$P \in \{P_{\pi_n}(d) | \forall n \in \mathcal{N}^+, \forall d \in [0,\varphi]\}. \tag{14c}$$

Moreover, the objective function (14a) is concave under the conditions in (14b). Shaked and Shanthikumar in [21] showed that the average waiting time of a $GI/GI/1$ queue with first-come-first-served order is jointly convex in the effective-arrival-rate and the service rate. The effective-arrival-rate of the corresponding coordinated queue (a $GI^{(m^*)}/D/1$ queue) of Q.2 is $\frac{P\lambda}{m}$. By the composition rule, we can see that $-h\left(\omega_{\pi_n}(d)\right)$ is jointly concave in $(\frac{P\lambda}{m}, d)$ (thus in (P,d)). This, together with the fact that $P\left(\frac{de^{-\beta d}}{\xi} - dp_e\right)$ is jointly concave in (P,d), implies that (14a) is a jointly concave function in (P,d). If we ignore the integer constraint in (14c), then Problem (14) can be solved efficiently by the gradient method, with the optimal solution denoted as (P^v, d^v). Accordingly, n^v can be obtained by solving (8) given (P^v, d^v). However, n^v obtained through this approach does not necessarily satisfy the integer constraint in (13b). In the following Lemma 3, we show that the optimal solution to Problem (13) can be easily obtained by rounding n^v to the nearest integer. In the lemma, we will use the notation of $d_n^* = \arg\max_d s(n,d)$. Then, we have the following characterization of the optimal solution (n^*, d^*) to Problem (13).

Lemma 3. *Given that (n^v, d^v) is an optimal solution to Problem (14a-b) (without considering the constraint (14c)), then the optimal solution to Problem (13) is either $(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor})$ or $(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$, whichever yields the larger objective function value.*[3]

Proof. First, we show that for any $\hat{n} < \lfloor n^v \rfloor$, $s(\hat{n}, d^*_{\hat{n}}) \le s(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor})$. It's equivalent to showing that for any $\hat{n} < \lfloor n^v \rfloor$, we can find an $(\lfloor n^v \rfloor, d_1)$ such that $s(\lfloor n^v \rfloor, d_1) \ge$

[3] $\lfloor n \rfloor$ and $\lceil n \rceil$ denote the largest integer no greater than n and the smallest integer no less than n.

$s(\hat{n}, d^*_{\hat{n}})$. From (6), $P_{\pi_n(d)}$ is monotonically increasing in both n and d. Thus, we can always find a point $(P_{\pi_{\lfloor n^v \rfloor}}(d_1), d_1)$ in the line segment between $(P_{\pi_{\hat{n}}}(d^*_{\hat{n}}), d^*_{\hat{n}})$ and (P^v, d^v). The monotonicity of $P_{\pi_n}(d)$ guarantees the existence and uniqueness of $(P_{\pi_{\lfloor n^v \rfloor}}(d_1), d_1)$. Due to the joint concavity of $\hat{s}$ in (P, d), we have $\hat{s}(P^v, d^v) \geq \hat{s}(P_{\pi_{\lfloor n^v \rfloor}}(d_1), d_1) \geq \hat{s}(P_{\pi_{\hat{n}}}(d^*_{\hat{n}}), d^*_{\hat{n}})$. Due to the equivalence between Problem (11) and Problem (14), we have $s(n^v, d^v) \geq s(\lfloor n^v \rfloor, d_1) \geq s(\hat{n}, d^*_{\hat{n}})$. Likewise, we can prove that for any $\hat{n} > \lceil n^v \rceil$, $s(\hat{n}, d^*_{\hat{n}}) \leq s(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$. Therefore, we can conclude that the optimal solution to Problem (13) is either $s(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor})$ or $s(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$. □

Lemma 3 indicates that we can obtain the optimal n^* by rounding n^v. What remains is how to calculate $d^*_{\lfloor n^v \rfloor}$ and $d^*_{\lceil n^v \rceil}$ efficiently. The following Lemma 4 indicates that $d^*_{\lfloor n^v \rfloor}$ and $d^*_{\lceil n^v \rceil}$ can be easily obtained using single-variable convex optimization methods, e.g., the gradient search method.

Lemma 4. *Given n, $s(n, d)$ is concave in d for $d \in \{d | \frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0\}$. Moreover, $s(n, d)$ is concave in d when $n = n^*$.*

Proof. We first prove that given any n, $s(n, d)$ is concave in d for $d \in \{d | \frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0\}$. For any d such that $\frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0$, $\frac{de^{-\beta d}}{\xi} - dp_e$ is a positive increasing concave function in d. Meanwhile, it can be seen from (5) that $P_{\pi_n}(d)$ is a positive decreasing concave function in d. Therefore, the product $P_{\pi_n}(d)\left(\frac{de^{-\beta d}}{\xi} - dp_e\right)$ is concave in d. According to Theorem 1, we have $\forall n \geq m$, $\frac{\partial^2 \omega_{\pi_n}(d)}{\partial d^2} > 0$, and $\frac{\partial \omega_{\pi_n}(d)}{\partial d} > 0$. This, together with the fact that $h(\omega)$ is a non-decreasing convex function, implies that $-h\left(\omega_{\pi_n}(d)\right)$ is also concave in d. Hence, $s(n, d)$ is concave in d for $d \in \{d | \frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0\}$.

We now prove that $s(n, d)$ is concave in d at an optimal n^*. This can be proved by showing that the condition $\frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0$ is satisfied at the optimal solution, which we will show by contradiction. Suppose that $\frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e < 0$ holds for an optimal solution (n^*, d^*). In this case, the objective in (13a) is monotonically decreasing in d, because the derivative of the first term in (13a) is negative in the domain and the second term in (13a) monotonically decreases with d. This contradicts with the assumption that (n^*, d^*) is an optimal solution. Thus $\frac{1}{\xi}\left(e^{-\beta d} - d\beta e^{-\beta d}\right) - p_e \geq 0$ must hold for an optimal solution to Problem (13). □

With Lemmas 3 and 4, we propose an optimal solution algorithm to Problem (13) in Fig. 6, which consists of the following steps.

- Step 1: Solve Problem (14) with gradient method. Calculate the corresponding n^v with (P^v, d^v).
- Step 2. a: Solve Problem (13) conditioning on $n = \lfloor n^v \rfloor$ with gradient method and denote the optimal solution as $(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor})$.
- Step 2. b: Solve Problem (13) conditioning on $n = \lceil n^v \rceil$ with gradient method and denote the optimal solution as $(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$.

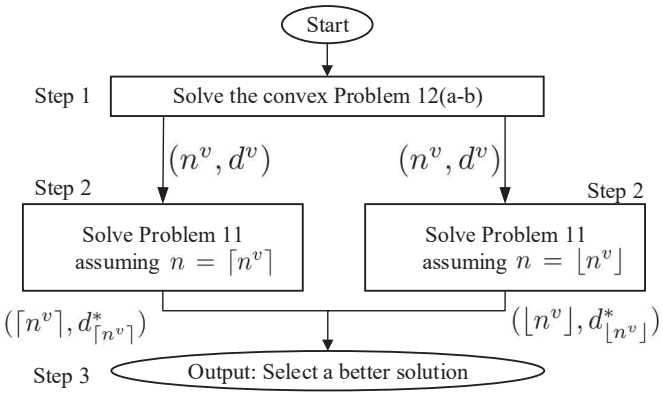


Fig. 6: Optimal Solution Algorithm Flowchart

- Step 3: If $s(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor}) > s(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$, return $(\lfloor n^v \rfloor, d^*_{\lfloor n^v \rfloor})$. Otherwise, return $(\lceil n^v \rceil, d^*_{\lceil n^v \rceil})$.

To recap, the optimal n and d obtained in this section defines the JoAP algorithm that optimizes the charging station operation. In particular, the charging station sets the charging price r that leads the EVs to request demand d. According to the current charging load at the station, the π_n admission policy decides whether to admit an EV or not. For those admitted EVs are admitted, the FIFO policy is applied to provide charging service.

V. Simulation Results

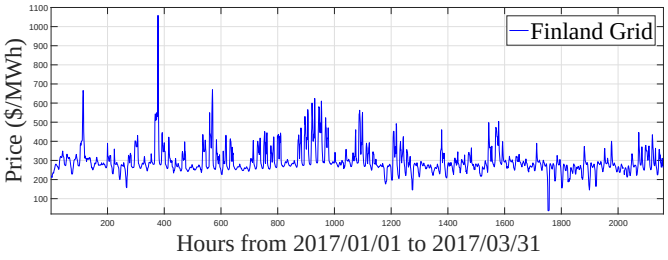
In this section, we use real-world data to evaluate the performance of the JoAP scheme and investigate how different system parameters affect the profit and admission performance. All the computations are solved in MATLAB [22] on a computer with an Intel Core i5-4670 3.40 GHz CPU and 8 GB of memory.

A. Experimental Setup

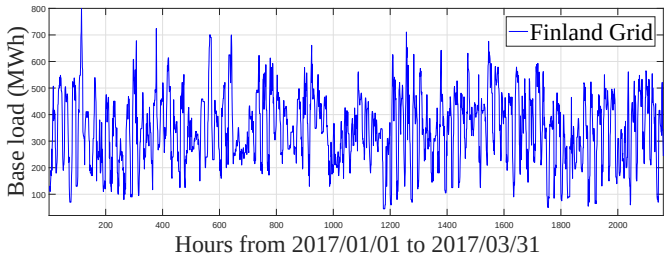
We base our simulations on the historical hourly electricity prices (Fig. 7a) and the base loads (Fig. 7b) of Finland Grid in the Nordic electricity market [23]. The data set spans the first 3 months of 2017. All the simulation results are the average performance of 90 days. The average prices and base loads are summarized in Fig. 7c. The EV arrivals follow a Poisson distribution, whose mean of different time periods are listed in Table. II. The settings of the peak hour match with the realistic vehicle trips in NHTS 2009 [24]. Unless specified otherwise, the charging station has $m = 4$ charging ports with a charging rate $\alpha = 11.5kW$. The number of parking lots is 40. All EVs have the same utility parameter $\beta = 0.05$ and the battery capacity $\varphi = 100kWh$.[4] For simplicity, we consider a linear waiting-time penalty $h(\omega) = c\omega$ [25], where $c > 0$ denotes the penalty rate. Our proposed JoAP algorithm is flexible enough to adapt its admission control and pricing methods to different EV arrival rates, penalty rates, and electricity prices.

TABLE II: Simulation historical data

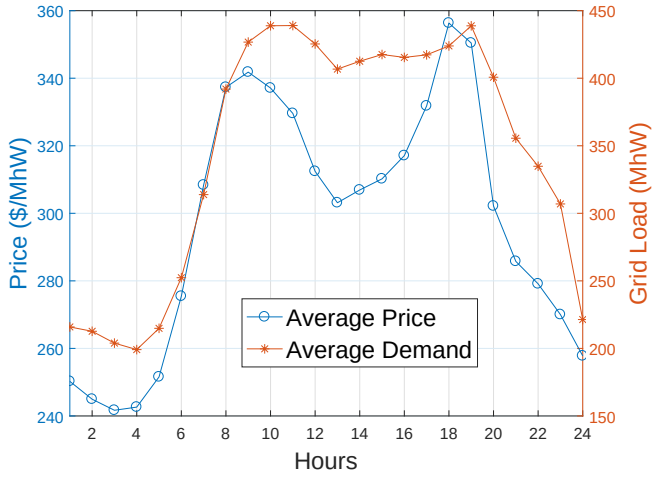
[4]The battery specifications follow the latest information from the Tesla website: https://www.tesla.com/models.



(a) The hourly ToU prices of Finland Grid



(b) The hourly base load of Finland Grid



(c) The average hourly prices and demand of Finland Grid

Fig. 7: Simulation Data from 2017/01/01 to 2017/03/31

Time of Day	08:01-12:00	12:01-16:00	16:01-20:00
λ (/minutes)	0.3	0.4	0.4
$\bar{p_e}$ ($/MWh)	330.2	309.3	335.2
Average Base (MWh)	432.5	413.1	320.2
Time of Day	20:01-24:00	00:01-04:00	04:01-08:00
λ (/minutes)	0.4	0.3	0.1
$\bar{p_e}$ ($/MWh)	273.2	244.9	293.2
Average Base (MWh)	304.7	207.9	293.3

For performance comparison, we consider the following two benchmark algorithms:

1) Queue-length based admission (QBA): An EV is admitted into the system only when the number of EVs already admitted is below a threshold. For our simulations, we set the threshold to the total number of parking lots in the charging station. Such an admission scheme has been widely used in current practice (e.g., California Plug-In Electric Vehicle Collaborative[5]).
2) Greedy admission: An EV is admitted if and only if doing so increases the system profit in the short-run (without considering future EV arrivals) [8].

B. Average Profit Evaluation

In Fig. 8, we compare the average profit per hour achieved by the three schemes under two different waiting-time penalty rates: $c = \$1/\text{min}$ and $c = \$0.4/\text{min}$. For each time period listed in Table I (scenarios), we simulate 1000 independent arrival processes $\mathcal{V}$ and plot the average profit performance.

We first compare the average profit of the entire day of three schemes. Fig. 8 shows that JoAP greatly outperforms the two benchmark schemes. The average profit over the whole day is 330% and 531% higher than that of the greedy admission scheme when the waiting-time penalty is low ($c = \$0.4/\text{min}$) and high ($c = \$1/\text{min}$), respectively. On the other hand, the widely used QBA scheme only achieves 44% of JoAP's average profit when waiting-time penalty is low, and a negative profit when waiting-time penalty is high.

Now we investigate the performance of the three schemes in different scenarios. During low-traffic period, e.g., from 4:01 to 8:00, the advantage of JoAP is not obvious. It only achieves 0.5% and 2% higher profit than the greedy algorithm under low and high waiting-time penalty rates, respectively. The advantage is more evident under heavy traffics, e.g., 12:01 to 24:00. Under the same traffic intensity, the advantage of JoAP over the greedy algorithm increases when p_e increases. This is because the admission probability decreases rapidly when p_e increases. On the other hand, the advantage of JoAP over QBA decreases when p_e increases. This is the profit of QBA is dominated by the delay penalty, and therefore is less sensitive to the increase of electricity price p_e.

It can be seen that the conventional QBA scheme performs very poorly with negative profit when the waiting-time penalty is high. In the events of bulky arrivals, the QBA scheme admits all EVs until there is no available parking lot and denies all the EVs that arrive later. This leads to heavy delay penalty for admitted EVs and high rejection rate for incoming EVs as well. The greedy admission scheme has a positive but low profit due to its inability to balance the charging schedule for the current and future EV arrivals. In fact, the greedy admission scheme always denies some EVs even under very light EV arrival traffic. In contrast, JoAP admits a proper number of EVs by jointly considering the EVs being served and the possible arrivals in the future, thus achieving a much higher profit than the two benchmark algorithms.

C. Admission probability Evaluation

In this subsection, we show that the average admission probability of JoAP scheme is comparable with that of the conventional QBA scheme. Fig. 9 compares the average admission probability of JoAP algorithm and the benchmarks under different penalty rates. Overall, the QBA scheme achieves the

[5]http://www.pevcollaborative.org/workplace-charging

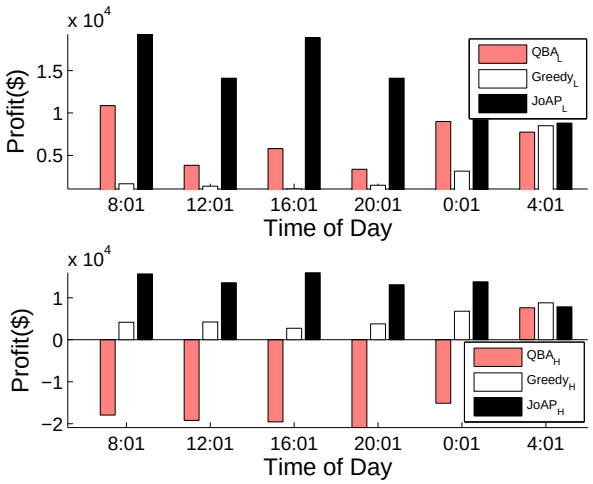


Fig. 8: Comparisons of the achieved profit of different methods in different time of a day. Figure above considers a low penalty rate ($c = \$0.4/\text{min}$), while figure below considers high penalty rate ($c = \$1/\text{min}$).

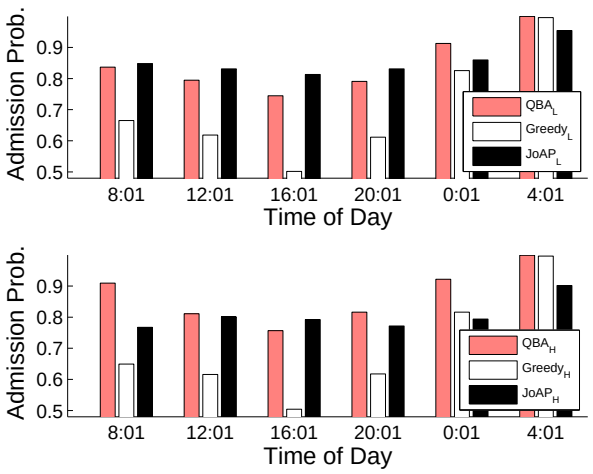
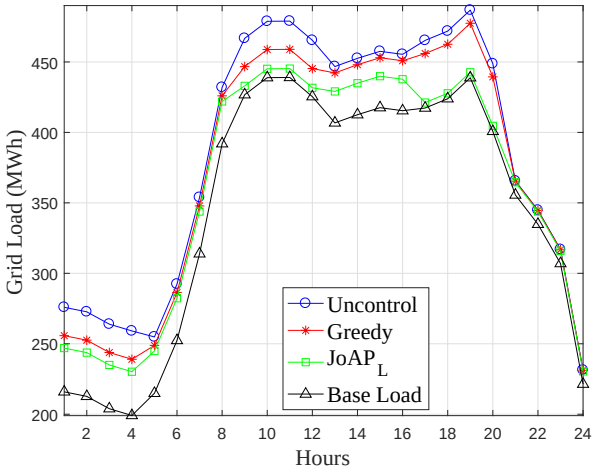
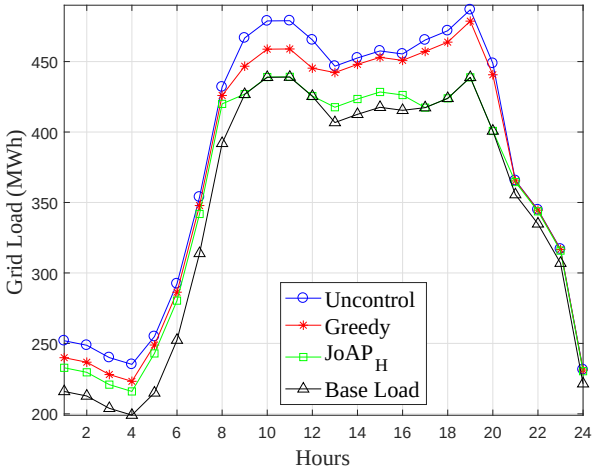


Fig. 9: Comparisons of the admission probability under a low penalty rate (figure above) and a high penalty rate (figure below).

highest admission probability, i.e., 86%, as it rejects an EV only when the parking lots are full. However, in the some periods with moderate arrival rates, e.g., 8:01 to 12:00, the admission probability of the QBA scheme falls below JoAP as it is oblivious to the possible future arrivals. The overall admission probability of the greedy admission algorithm is the lowest, i.e., 70% and 69% in the light-penalty-rate and high-penalty rate cases, respectively. JoAP algorithm has an admission probability 85% and 80% in the light-penalty-rate and high-penalty rate cases, and achieves a good balance between high admission probability and high profit. Because the conventional QBA scheme achieves negative profit and the superior in the admission probability is limited, we consider only the Greedy method and JoAP in the following simulations.



(a) low penalty rate



(b) high penalty rate

Fig. 10: Comparisons of the average total grid load under different penalty rates

D. *Impact to the Grid*

In this subsection, we show that JoAP, by responding to the ToU price, effectively reduces the peak load demand in the power grid. In Fig. 10, we plot the total grid load, i.e., the base load plus the charging load, under JoAP, the Greedy method, and the uncontrolled charging scheme, respectively. The uncontrolled charging scheme charges each EV independently with the EV-surplus-maximum demand immediately after it arrives. For each day, we simulate 100 independent arrival processes $\mathcal{V}$ and plot the average grid load over 90 days. Compared with the uncontrolled charging scheme, JoAP reduces the average peak load by 9.5% and 10.9% under the low and high penalty rates, respectively. In comparison, the Greedy method only reduces the average peak load by 1.9% and 2.1%, respectively. This is because, when the base load is heavy, the ToU electricity price p_e is also high with high probability (see Fig. 7c). In response to high p_e, JoAP would raise the charging service price and admit fewer EVs. As a result, the total charging load is reduced. On the other hand,

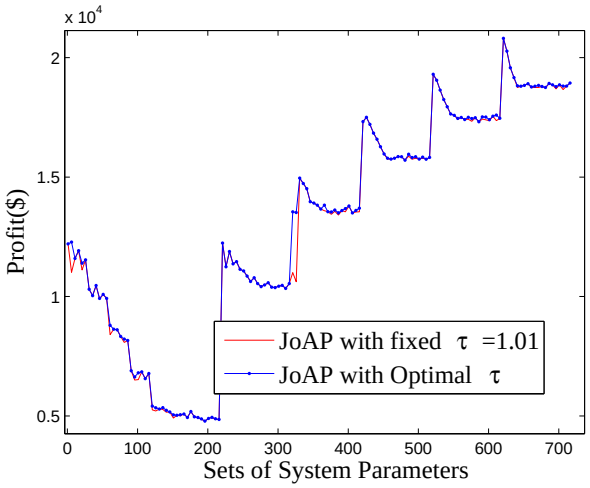


Fig. 11: The achieved average profit with $\tau = 1.01$ and the optimized τ comparison versus (c, λ, p_e)

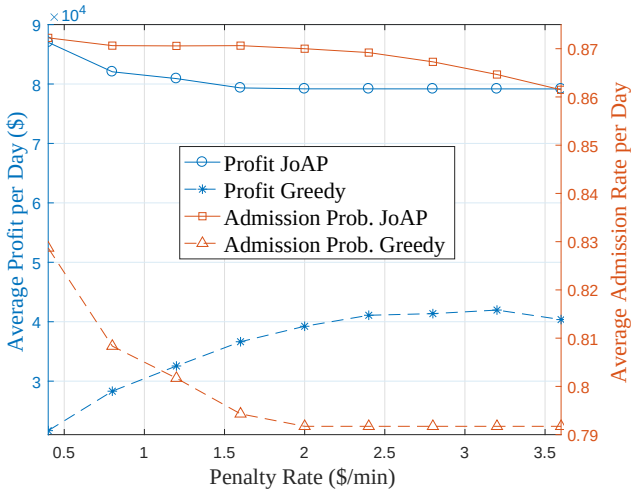


Fig. 12: The achieved average profit and admission probability comparison versus c

the Greedy algorithm is less sensitive to p_e. Notice that a low peak load demand is critical to the safety, stability, and economic viability of power grids. Thus, JoAP benefits not only the charging station owners, but the power grid as a whole.

E. Impact of τ and c

1) Impact of τ: We have considered a fixed τ in the theoretical analysis in Section III.B. In Fig. 11, we numerically evaluate the performance gain if we optimize the value of τ, and comparing with the case of using a fixed value of $\tau = 1.01$. For each (c, λ, p_e), we simulate 100 independent arrival processes $\mathcal{V}$ and plot the average profit performances with $\tau = 1.01$ and the optimized τ. Averaging over all scenario, *optimizing over τ increases the profit over fixing $\tau = 1.01$ only by* 5.9%. Therefore, we can focus on the optimizing of n (with a fixed τ) in practice.

2) Impact of c: We now turn to investigate how different waiting penalty rate impacts the admission and pricing under JoAP and the Greedy method. For each c, we simulate 100 independent arrival processes $\mathcal{V}$ for each day. Take average over 90 days, we plot the average daily profit and admission probability versus the penalty rate c in Fig. 12. On one hand, we observe that the profit and the admission probability of JoAP decrease slightly, i.e., by 8.9% and 1.3%, when increasing the penalty rate from 0.4 to 3.6 \$/min. On the other hand, we observe that the profit of the Greedy method increases by 98.5% when increasing the penalty rate from 0.4 to 3.2 \$/min and decreases by 6.6% when increasing the penalty rate from 3.2 to 3.6 \$/min. More importantly, we observe that the admission probability of the Greedy method decreases rapidly when the penalty rate is increased from 0.4 to 2 \$/min. Overall, the average profit of JoAP over all the penalty rates is 118% higher than that of the greedy admission scheme. This implies that JoAP is flexible enough to adapt its admission control and pricing methods to different EV arrival rates, penalty rates, and electricity prices. Meanwhile, the Greedy method may suffer instability when the penalty rate changes.

VI. Conclusions

In this paper, we proposed a novel joint admission and pricing (JoAP) mechanism for a EV charging station to maximize its profit. In contrast to existing EV charging operation schemes, the JoAP scheme adopts a multi-sub-process admission control policy that is capable of balancing between the system admission probability and the EVs' QoS requirements according to the EV arrival rate, the electricity price, and the delay penalty. We introduced a tandem queueing model to analyze the joint admission control and scheduling process, and proposed an efficient algorithm to compute the optimal solution. Simulation results showed that JoAP can effectively increase the charging station's profit while providing good QoS guarantees to the EV users and reducing the load in the power grid.

In our future study, we plan to extend this work to the more general case with multi-class EVs. We will further consider how the integration of renewable, Vehicle to Grid services, and distributed energy generations will impact the admission control and efficiency of the charging station.

Appendix A
Proof of Lemma 1

Proof. The mentioned two-phase-process approximation is replacing each server process with deterministic service time T_v by two-phase process with exponential distribution with rate $\kappa = \frac{2}{T_v}$ (Fig. 13). In particular, take the Laplace transform of the service distribution, we have $f^*(s) = \frac{\kappa r_1 s + \kappa^2(r_1 + r_2 - r_1 r_2)}{s^2 + 2\kappa s + \kappa^2(r_1 + r_2 - r_1 r_2)}$. By simple algebra, we derive that $\kappa = \frac{2}{T_v}$, $r_1 = -1$, and $r_2 = \frac{5}{4}$ when the service time is deterministic and equals the constant T_v [16]. Upon this approximation, we denote the system state as a two-dimension pair (s_1, s_2), where s_1 is the number of busy processes in phase 1, and $s_2 - s_1$ is the number of busy processes in

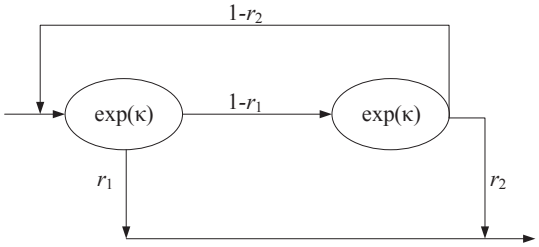


Fig. 13: Two-phase process

phase 2. For a particular state (s_1, s_2), it can transfer to at most 6 neighborhood states: (s_1-1, s_2), (s_1, s_2), (s_1+1, s_2), (s_1-1, s_2-1), (s_1, s_2-1), and (s_1+1, s_2+1). Let $\boldsymbol{T}^k$ denote the generator matrix when there are total k servers. Then, we calculate generator matrix $\boldsymbol{T}^k$ in following 4 cases.

1) If $s_1 \geq 1$ and $s_2 < k$, the system can transfer to all 6 states mentioned above. The non-zero elements in $\boldsymbol{T}^k$ are,
   a) $T^k_{(s_1,s_2),(s_1-1,s_2)} = s_1(1-r_1)\kappa$;
   b) $T^k_{(s_1,s_2),(s_1+1,s_2+1)} = \lambda$;
   c) $T^k_{(s_1,s_2),(s_1-1,s_2-1)} = s_1 r_1 \kappa$;
   d) $T^k_{(s_1,s_2),(s_1+1,s_2)} = (s_2 - s_1)(1-r_2)\kappa$;
   e) $T^k_{(s_1,s_2),(s_1,s_2-1)} = (s_2 - s_1) r_2 \kappa$;
   f) $T^k_{(s_1,s_2),(s_1,s_2)} = -s_2\kappa - \lambda$.
2) If $s_1 = 0$ and $s_2 < k$, the system can transfer to (s_1, s_2), (s_1+1, s_2), (s_1, s_2-1), (s_1+1, s_2+1).
3) If $s_1 = 0$ and $s_2 = k$, the system can transfer to (s_1, s_2), (s_1+1, s_2), (s_1, s_2-1).
4) If $s_1 = 0$ and $s_2 = 0$, the system can transfer to (s_1, s_2), (s_1+1, s_2+1).

After manipulation and observation, we have

$$\mathbf{T}^{k+1} = \begin{pmatrix} \mathbf{T}^k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{k+1} \end{pmatrix}, \tag{15}$$

where $\mathbf{T}^{k+1}$ and $\mathbf{T}^k$ are the generate matrix for $n = k+1$ and $n = k$, respectively.

Let $\mathbf{x^{k+1}} = \begin{pmatrix} \mathbf{y^{k+1}} \\ \mathbf{z^{k+1}} \end{pmatrix}$ and $\mathbf{x^k}$ denote the steady-state probability for $\boldsymbol{T^{K+1}}$ and $\boldsymbol{T}^k$, respectively. The steady-state probability of state (s_1, s_2) is denoted as $x^k_{s_1(k-1)+s_2}$. Substitute $\mathbf{x^k T_k} = 0$ and $\mathbf{x^{k+1} T_{k+1}} = 0$ into equation (15), we have follows,

$$\mathbf{y^{k+1}} = \xi \mathbf{x^k}, \tag{16a}$$

$$\mathbf{x^{k+1}_{(j,k)}} = \binom{k}{j}(1-r_1)^j \mathbf{x^{k+1}_{(0,k)}}, \quad \forall j \in \{1, 2, ..., k\}, \tag{16b}$$

$$\mathbf{x^{k+1}_{(j,k+1)}} = \binom{k+1}{j}(1-r_1)^j \mathbf{x^{k+1}_{(0,k+1)}}, \tag{16c}$$

$$\forall j \in \{1, 2, ..., k+1\}, \tag{16d}$$

$$\mathbf{x^{k+1}_{(0,k+1)}} = \frac{d_v \lambda}{3(n+1)} \mathbf{x^{k+1}_{(0,k)}}, \tag{16e}$$

where ξ is a scalar. Let $P_k(k+1, d) = \sum_{j=0}^{k} \mathbf{x^{k+1}_{(j,k)}}$ denote the steady-state probability of $s_2 = k$. Substitute equation (15) into $P_k(k+1, d) = \sum_{j=0}^{k} \mathbf{x^{k+1}_{(j,k)}}$ and $P_{k+1}(k+1, d) = \sum_{j=0}^{k+1} \mathbf{x^{k+1}_{(j,k+1)}}$. After manipulation, we get

$$P_{k+1}(k+1, d) = \frac{\tau n d \lambda}{m} P_k(k+1, d). \tag{17}$$

With boundary condition $\sum_{j=0}^{n} P_j(n, d) = 1$, we get the steady-state distribution from equation (17),

$$P_j(n, d) = \frac{\frac{d^j \eta^j}{j!}}{\sum_{j'=0}^{n} \frac{\eta^{j'}}{j'!}}, \text{where} \quad \eta = \frac{\tau n \lambda}{m}. \tag{18}$$

□

REFERENCES

[1] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "Optimal planning of PEV charging station with single output multiple cables charging spots," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 1–10, Jan 2016.

[2] A. Schroeder and T. Traber, "The economics of fast charging infrastructure for electric vehicles," *Energy Policy*, vol. 43, pp. 136 – 144, Apr 2012.

[3] P. You and Z. Yang, "Efficient optimal scheduling of charging station with multiple electric vehicles via V2V," in *IEEE International Conf. Smart Grid Commun.(SmartGridComm)*, Venice, Italy, pp. 716–721, Nov 2014.

[4] W. Tang and Y. J. Zhang, "A model predictive control approach for low-complexity electric vehicle charging scheduling: Optimality and scalability," *IEEE Trans Power Sys.*, vol. 32, no. 2, pp. 1050–1063, March 2017.

[5] W. Tang, S. Bi, and Y. J. Zhang, "Online coordinated charging decision algorithm for electric vehicles without future information," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2810–2824, Nov 2014.

[6] L. Zhang and Y. Li, "A game-theoretic approach to optimal scheduling of parking-lot electric vehicle charging," *IEEE Trans. on Veh. Technol.*, vol. 65, no. 6, pp. 4068–4078, 2016.

[7] W. Yuan, J. Huang, and Y. J. Zhang, "Competitive charging station pricing for plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 1–13, Dec 2015.

[8] Z. Wei, J. He, and L. Cai, "Admission control and scheduling for EV charging station considering time-of-use pricing," in *Vehicular Technology Conf.(VTC Spring)*, Nanjing, China pp. 1–5, May 2016.

[9] Y. Kim, J. Kwak, and S. Chong, "Dynamic pricing, scheduling, and energy management for profit maximization in phev charging stations," *IEEE Trans. on Veh. Technol.*, vol. 66, no. 2, pp. 1011–1026, 2017.

[10] J. Erickson, N. Guan, and S. Baruah, "Tardiness bounds for global EDF with deadlines different from periods," in *Principles of Distributed Systems (OPODIS)*, Tozeur, Tunisia, pp. 286–301, 2010.

[11] M. A. Khan and U. Toseef, "User utility function as quality of experience (QoE)," in *International Conf., Networks (ICN)*, Valetta Malta, pp. 99–104, Jan 2011.

[12] H. H. Gossen, *The laws of human relations and the rules of human action derived therefrom*. Mit Press, 1983.

[13] C.-p. Li and M. J. Neely, "Delay and power-optimal control in multi-class queueing systems," submitted for publication, available on-line at https://arxiv.org/abs/1101.2478.

[14] M. Vuuren, I. J. Adan, and S. A. Resing-Sassen, "Performance analysis of multi-server tandem queues with finite buffers and blocking," in *OR Spectrum*, vol. 27, no. 2, pp. 315–338, 2005.

[15] A. O. Allen, *Probability, statistics, and queueing theory*. Academic Press, 2014.

[16] H. C. Tijms, "Note on approximations for the multiserver queue with finite buffer and deterministic services," *Probability in the Engineering and Informational Sciences*, vol. 22, no. 4, pp. 653–658, Sept 2008.

[17] J. Sztrik, *Basic queueing theory.* University of Debrecen, , 2012.

[18] H. Tijms, "New and old results for the m/d/c queue," *AEU-International Journal of Electronics and Communications*, vol. 60, no. 2, pp. 125–130, March 2006.

[19] H. C. Tijms, *A first course in stochastic models*. John Wiley & Sons, Inc., 2003.

[20] J. Kingman and M. Atiyah, "The single server queue in heavy traffic," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 57, no. 4, p. 902–904, 1961.
[21] M. Shaked and J. G. Shanthikumar, "Stochastic convexity and its applications," *Advances in Applied Probability*, vol. 20, no. 2, pp. 427–446, 1988.
[22] MATLAB, *version R2016b*. Natick, Massachusetts: The MathWorks Inc., 2016.
[23] Nordpoolspot.com, "Historical market data," 2017. [Online]. Available: http://www.nordpoolspot.com/historical-market-data/. [Accessed: 23- Aug- 2017].
[24] A. Santos, N. McGuckin, H. Y. Nakamoto, D. Gray, and S. Liss, "Summary of travel trends: 2009 national household travel survey," tech. rep., 2011.
[25] C. S. Yeo and R. Buyya, "Service level agreement based allocation of cluster resources: Handling penalty to enhance utility," in *IEEE Int. Cluster Comput. (Cluster)*, Burlington, MA, USA, pp. 1–10, Sept 2005.