Capacity Planning Frameworks for Electric Vehicle Charging Stations With Multiclass Customers

Islam Safak Bayram, Member, IEEE, Ali Tajer, Member, IEEE, Mohamed Abdallah, Senior Member, IEEE, and Khalid Qaraqe, Senior Member, IEEE

Abstract—In order to foster electric vehicle (EV) adoption, there is a strong need for designing and developing charging stations that can accommodate different customer classes distinguished by their charging preferences, needs, and technologies. By growing such charging station networks, the power grid becomes more congested and, therefore, controlling charging requests should be carefully aligned with the available resources. This paper focuses on an EV charging network equipped with different charging technologies and proposes two frameworks. In the first framework, appropriate for large networks, the EV population is expected to constitute a sizable portion of the light duty fleets. This necessitates controlling the EV charging operations to prevent potential grid failures and distribute the resources efficiently. This framework leverages pricing dynamics in order to control the EV customer request rates and to provide a charging service with the best level of quality of service (QoS). The second framework, on the other hand, is more appropriate for smaller networks, in which the objective is to compute the minimum amount of resources required to provide certain levels of QoS to each class. The results show that the proposed frameworks ensure grid reliability and lead to significant savings in capacity planning.

Index Terms—Capacity planning, distributed control, electric vehicles (EVs), performance evaluation, pricing.

NOMENCLATURE

\[ \beta_j(k) \] Loss-of-load-probability (LoLP) for class \( j \).
\[ \delta_j \] LoLP target for customer class \( j \).
\[ \lambda_j(k) \] Aggregated arrival rate for customer class \( j \) at time \( k \) given in (1).
\[ \lambda_j \] Arrival rate for customer class \( j \) at time \( k \).
\[ b_j \] Amount of resources requested by customer class \( j \).
\[ C_k \] Amount of power drawn from the grid during period \( k \).

I. INTRODUCTION

ELECTRIC vehicles (EVs) are becoming a viable transportation option as they offer solutions to an array of current societal problems ranging from high oil prices to environmental concerns. Corollary, EV population is expected to reach a sizable market portion in the next decade (e.g., 10% of the U.S. national fleet and similar targets in Europe) [1]. However, achieving such penetration rates requires wide deployment of charging facilities that can serve different types of charging requests. Therefore, if not controlled, EV charging can easily lead to transformer and line overloading and it can deteriorate the power quality and even endangers the security of supply. The impacts of EV charging on the grid has been well documented in [1]–[4]. For instance, Hadley [4] argues that if 5% of the EVs charge simultaneously using fast charging technology, 5.5 GW of extra power would be needed in Virginia and Carolinas region by 2018. Similarly, North American Electric Reliability Corporation regions would require an expansion of 5.5% in their power generation capacity in a typical EV penetration rate of 25%.

As a result, distribution grid could easily become a bottleneck. For example, Bayram et al. [1] stated that even adding two level-2 chargers in a typical neighborhood in the U.S. could easily cause transformer overloading. Also, uncontrolled demand can decrease the efficiency of the power grid operations and increase power generation cost if it occurs during peak hours.

\[ p_j(k) \] Price of service for customer type \( j \) at time \( k \).
\[ Q_C^j \] Number of simultaneous customers of type \( j \) requesting \( b_j \) units.
\[ q_j \] Average number of arrival rates of class \( j \).
\[ R \] Aggregated utility of all customers.
\[ S \] Aggregated customer load given in (12).
\[ U^n(\cdot) \] Utility function for customer \( n \).
\[ \lambda^a \] Arrival rate vector for all customer classes at period \( n \).

II. FRAMEWORKS DESIGN

A. Motivation

MANUSCRIPT received April 22, 2014; revised August 23, 2014 and December 8, 2014; accepted February 4, 2015. Date of publication March 13, 2015; date of current version June 18, 2015. This work was supported by the National Priorities Research Program through the Qatar National Research Fund (a member of Qatar Foundation) under Grant 6-149-2-058. Paper no. TSG-00337-2014.

I. S. Bayram is with the Qatar Environment and Energy Research Institute, Qatar Foundation, Doha 5825, Qatar (e-mail: ibayram@qf.org.qa).

A. Tajer is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA.

M. Abdallah is with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha 23874, Qatar, and also with the Department of Electronics and Communications Engineering, Cairo University, Giza 12613, Egypt.

K. Qaraqe is with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha 23874, Qatar.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2015.2406532
B. Related Works

There exists an increasing body of literature on developing intelligent charging station architectures [5]–[11], controlling and scheduling EV demand [1], [12]–[16], and rather limited literature on capacity planning on charging infrastructures serving multiple classes of customers [17], [18]. This section provides a brief overview of related literature.

The studies on charging station design can be divided into two categories. The first category includes works from power engineering perspective [5]–[7], in which the studies aim to minimize the charging duration by improving the efficiency of power electronics and equipping the system with energy storage unit. On the other hand, approaches in the second category are mostly focused on the system level where they abstract the details of the underlying power system components and evaluate the system performance in terms of long-term statistical metrics, e.g., mean waiting time and percentage of served customer using the arguments from queuing theory [8], [9], [11]. Bayram et al. [8] presented a small scale fast charging station design and blocking probability, that is the probability request of an EV is rejected, is used as the main performance metric to solve the optimal resource provisioning problem. Sungwoo and Kwasinski [10] used M/M/s queues to model the EV demand at fast charging stations near highway exits. In [9], residential charging infrastructures are modeled using M/M/∞ queue, where they broadcast the arrival rates to customers such that probability of blocking (exceeding circuit capacity) is zero. Furthermore, Liang et al. [11] used queueing networks model to estimate the EV charging demand. In this paper, we perform a more holistic approach by taking into account that the charging station can serve multiple classes of customer which are differentiated by the charger technology used. Also, we employ LoLP as our main performance metric, which measures the probability that grid resources cannot accommodate EV demand [19].

The pricing-based control methods have been employed successfully in communication networks where the goal is to match scarce bandwidth resources to insatiable user demand [20]. In a similar fashion, pricing mechanism applications gained popularity in the literature on controlling EV demand [1], [13]–[16], [21]. References [1] and [16] proposed a control mechanism in a network of charging stations to route customers to idle stations. Xi and Sioshansi [13] proposed a pricing scheme for EV charging which leads to socially optimal solution, whereas Fan [21] used proportional fairness pricing from communication literature to control the charging rates of EVs. In this paper, we also consider the technological constraints and current state of affairs that charging rate is constant until the end of the service. Reference [15] is most relevant to this paper, which models general power consumers behavior with utility functions and propose a pricing algorithm to control the consumption of smart grid users. This paper mainly focuses on multiclass EV chargings and our framework ensures grid reliability at all times. The literature on resource provisioning is rather limited. Bayram et al. [17] presented a capacity planning framework in a large-scale charging stations with single class customers. This framework is based on computing a deterministic quantity effective power using on-off Markov models. In this paper, we assume that charging infrastructure can serve multiple types of customers.

The proposed methodology is rooted in multidimensional loss systems in teletraffic engineering, where the goal is to provide statistical quality of service (QoS) guarantees to customers with different demand profiles. The control mechanisms with congestion pricing in multirate Erlang-B systems and related resource provisioning problems are addressed in [22]–[26]. References [21] and [23]–[25] focused on efficient computation of blocking probabilities and their derivatives, whereas [22] and [26] provided efficient algorithms to solve bandwidth provisioning problem in congestible networks.

II. Proposed Frameworks

EV charging demands are primarily dominated by customers’ needs and preferences. Based on this premise, designing different aspects of the network of charging stations is governed by the economic dynamics between network operators, on one hand, and the customers on the other hand. In such dynamics, the former seek operational reliability and profit margins and the latter seek economic incentives. Such dynamics and their pertinent design challenges vary by the size of network and customers population. In this paper, we consider two canonical design problems for EV charging infrastructures fed by a single substation suited for large metropolitan areas and small cities as summarized in Fig. 1.

Specifically, the first framework proposes a pricing-based EV control for charging stations located in large metropolitan cities in which the EV population is assumed to be large. In this framework, the objective is to incentivize and control the customers to submit their charging demands at certain optimal rates which maximize a social welfare utility. This pricing-based approach framework, at its core, enables station operators to alleviate the congestion and the degradations in the service quality during increased demand periods through incentivizing the customers to defer their charging needs to less congested periods. This control mechanism ensures charging services with desirable level of QoS guarantees.

The second framework focuses on small cities, in which customer demands can be obtained from profiling studies, and proposes a capacity provisioning mechanism for EV
charging stations. Contrary to large metropolitan areas, small cities with fewer charging stations and customers enjoy less flexibility for incentivizing the customers. Hence, the primary goal in this framework is to compute the minimum amount of required resources to provide charging service to meet a target level of QoS.

A. Network Model

We consider a network of charging stations serving $N$ customers. Since the customers do not necessarily seek identical charging services, we consider $J$ distinct customer classes, which are distinguished by their preferences; sizes of their battery packs, amounts of requested demands, and the available charger technologies. The charging rate of customers of type $j \in \{1, \ldots, J\}$ is denoted by $b_j$. Since the amount of resources are finite, upon the arrival of a specific customer type $j$, if the available resources is less than $b_j$ the customer will not be served, resulting in an outage. Hence, the probability of being blocked (or LoLP) constitutes a natural system performance metric. Due to the temporal fluctuation network conditions (e.g., congestion), the amount of available resources varies over time. Hence, we consider a dynamic system indexed by the time index $k \in \mathbb{N}$, in which $C_k$ denotes the aggregate units of grid power available to the entire EV fleet at time $k$. Note that $C_k$ is an exogenous parameter given to station. Since the charging stations reside in a small well-confined regions, the set of chargers can be abstracted collectively as a super-station with multiple classes of customers. We further assume that system capacity is large when compared to demands of each EV type.

Increased outage events lead to reduced utility for the network operator and service disruption for the customers. Hence, optimizing the operation of the network strongly hinges on finding the decision rules that lead to the optimal level of LoLPs that is defined as the probability that grid resources fall short of aggregated customer demand [19]. The leverage that the network operators have for adjusting the LoLP rates at the desired optimal rates is service pricing, by using which they can influence the behavior of the customers abstracted by their arrival rates. To formalize this, we define $p_j(k)$ as the price of service of type $j$ and define the price vector $\mathbf{p}(k) \triangleq [p_1(k), \ldots, p_J(k)]$, accordingly.

Based on the prices $\mathbf{p}(k)$ and its needs, each customer decides whether to generate a service request of a certain type. We denote the rate of type $j \in \{1, \ldots, J\}$ requests generated by customer $n \in \{1, \ldots, N\}$ during period $k$ by $\lambda^n_j(k; \mathbf{p}(k))$. Hence, the aggregate arrival rate of service requests of type $j$ is

$$\lambda_j(k; \mathbf{p}(k)) \triangleq \sum_{n=1}^{N} \lambda^n_j(k; \mathbf{p}(k)). \quad (1)$$

From the network operators point of view, these requests are being constantly generated across the networks and as discussed in [1], [9], and [10], a common assumption is that such aggregated requests (and not necessarily individual requests) arrive according to a Poisson process. Accordingly, we define the arrival rate vectors

$$\lambda^n_j(k; \mathbf{p}(k)) \triangleq [\lambda^n_1(k; \mathbf{p}(k)), \ldots, \lambda^n_J(k; \mathbf{p}(k))] \quad (2)$$

and

$$\lambda(k; \mathbf{p}(k)) \triangleq [\lambda_1(k; \mathbf{p}(k)), \ldots, \lambda_J(k; \mathbf{p}(k))]. \quad (3)$$

By noting that the LoLPs for each class are functions of the arrival rates, finally, we define $\beta_j(k; \lambda(k; \mathbf{p}(k)))$ as the probability that customers of type $j$ are blocked and define $\beta(k; \lambda(k; \mathbf{p}(k)))$ as the corresponding LoLP vector. Since different customer types do not necessarily charge their batteries at the same rate, we denote the average charging duration of customers of type $j$ by $1/\mu_j$. It is noteworthy that from station operator standpoint, the information required to characterize customers’ profiles are the average service duration and the charging rates. This is depicted in Fig. 3. These parameters enable modeling customers’ behaviors with fine granularity.

The main goal is to keep the aggregated demand below a constant power $C_k$ with minimal or controlled loss-of-load (outage) events. Drawing constant power is highly desirable from the viewpoint of the network operators and its benefits include.

1) Grid components are isolated from stochastic variations and hence grid reliability is ensured [1], [8].

2) Station operator can establish long-term contracts and benefit from the lower prices.

3) Constant demand reduces the peak-to-average demand ratio of the whole power system and accordingly the average spot prices would reduce.

4) It leads to more efficient market equilibrium [1].

It is noteworthy that since the charging stations reside in a small well-confined region, the whole set of chargers act as one big station with multiple classes of customers. Due to this fact, power system losses are assumed to be negligible and all customers share the same resource pool.

For the simplicity of the notations, in the rest of this paper, the explicit dependence on $k$ is omitted.

B. Large Networks: Pricing-Based Control Framework

As the EV population is expected to constitute a sizable portion of the national light duty fleets, there is a need to control EV charging operations in order to protect the grid components from potential failures. To this end, this framework focuses on charging infrastructures that are located in large metropolitan areas with high EV penetration rates. The primary goal is to leverage pricing in order to control the EV customer arrival rates such that station operator can provide a charging service with a small level of customer LoLP QoS for each class. Here, the primary goal is to control the large EV demand in a large city for a given amount of grid power $(C_k)$, which is assumed to be computed by the utility operator and meets the power network constraints (e.g., congestion and transformer ratings). Pricing-based control scheme is depicted in Fig. 2. Specifically, the goal is to design a distributed control framework for maximizing an aggregate utility (social welfare measure), which requires that the EVs operate
at an optimal set of arrival rates. In order to establish the tools for the distributed design, we define utility function
\[ U^n(\lambda^n; \beta(\lambda)) \]
for a given set of arrival rates \( \lambda^n \) and block probabilities \( \beta \). Hence, the aggregate utility in the network is
\[ R \triangleq \sum_{n=1}^{N} U^n(\lambda^n; \beta(\lambda)). \tag{4} \]
Therefore, the social welfare problem can be defined as the problem that maximizes the aggregate utility \( R \) over all possible choices of the arrival rates \( \{\lambda^n\}_{n=1}^{N} \), that is
\[ \max_{\{\lambda^n\}_{n=1}^{N}} \sum_{n=1}^{N} U^n(\lambda^n; \beta(\lambda)). \tag{5} \]
We assume the utility function \( U^n \), is increasing in the arrival rates \( \{\lambda_j\}_{j=1}^{J} \), and decreasing in the LoLPs \( \{\beta_j\}_{j=1}^{J} \). Furthermore, it is assumed that \( U^n \) is concave in \( \{\lambda_j\}_{j=1}^{J} \) and continuously differentiable in all of its arguments. This problem is treated in Section III.

C. Small Networks: Resource Provisioning Framework

In the resource provisioning problem, we are interested in computing the minimum amount of grid resources \( C \) such that system operator guarantees certain levels of reliability for serving different types of users by enforcing \( \beta_j \leq \delta_j \) for each customer type. This structure suits the resource provisioning problems in small cities in which customer arrival rates can be obtained from profiling studies with reliable accuracy. Based on the constraints defined for LoLPs, the problem can be cast as
\[ C^n = \begin{cases} \min C \\ \text{s.t. } \beta_j \leq \delta_j, \text{ for } j \in \{1, \ldots, J\}. \end{cases} \tag{6} \]
This problem is treated in Section IV.

D. Toy Example

We provide a toy example in order to put some of the details into perspective. Even though the objectives of the proposed frameworks are different, computation of LoLP is carried out for a fixed parameter setting (e.g., \( \lambda \) and \( \mu \)), the methods for computing blocking probability is valid for both cases and details are given in Section V. Let us assume that the charging infrastructure draws \( C = 1000 \) units of power from the grid and serves three types of EV customers, i.e., \( j \in \{1, 2, 3\} \). Customer classes are differentiated by the charging technology they use, and by mimicking the current charging standards (fast charging, level-II three, and single phase). It is assumed that \( b = \{50, 7, 5\} \), the charging rates are \( \mu_1 = 3 \), \( \mu_2 = 0.42 \), and \( \mu_3 = 0.2 \), and the arrival rates for each type are \( \lambda_1 = \lambda_2 = \lambda_3 = 14 \) (on average 14 customers arrive at each time unit). For this specific simulation setting, the resulting LoLPs are \( \beta^1_{1000} = 0.0152 \), \( \beta^2_{1000} = 0.0015 \), and \( \beta^3_{1000} = 0.0011 \). Note that customers of type 1 have the highest LoLP mainly because grid resources are shared equally among all customer classes and type 1 customers use fast charging technology which draws more power compared to other classes. Next, we evaluate the same system for a wide range of arrival rates. We assume that the sum \( \lambda = \sum_{j=1}^{3} \lambda_j \) varies in the range \([1, 80]\) and \( \lambda_j = \lambda/3 \) for \( j \in \{1, 2, 3\} \). The results are depicted in Fig. 4. It is important to notice that, for the given amount of grid resources (in our case grid allows \( C = 1000 \)), stations can provide a very good service for light traffic scenarios in which the aggregated arrival rate varies in the range \( \lambda \in [1, 50] \), which can be a very typical setting for small cities. It is noteworthy that instead of providing an extremely stringent LoLP constraints (near zero outage probability), network operator can back off from \( C = 1000 \) to a smaller level \( C' < C \) while guaranteeing a high level of LoLP (e.g., 1% LoLP) for each customer class.

On the other hand, in the heavy traffic regimes (e.g., \( \lambda > 50 \)) which can occur in large metropolitan cities, in order to provide good QoS, either charging resources (\( C \)) should be increased or the arrival rates should be controlled. To shed more light on this, we evaluate the system performance for a fixed set of arrival rates \( \{\lambda_1 = \lambda_2 = \lambda_3 = 10\} \) and for a range of serving capacity values \( C \in [500, 1000] \). As clearly depicted in Fig. 5, more customers will be accommodated as the amount of resources increases. Hence, besides upgrading...
the grid to have increased capacity, which might not be economically viable in the short term, an alternative approach is to provide charging services based on the available resources and with the best possible QoS via controlling customer arrival rates. The pricing-based control framework discussed in Section III-C attacks this problem.

III. PRICING-BASED CONTROL

A. Global Problem

The treatment provided in this section has a twofold purpose. Primarily, it provides an optimal LoLP policy in order to maintain an optimal social welfare utility as characterized in (4). Secondly, it proves that the solution provided is amenable to distributed implementation. This is of paramount significance as in an EV charging network the EVs are autonomous entities and operate based on the common information known to all the EVs (e.g., charging prices) and their local perception about network dynamics. Such distributed nature of the EV networks necessitates solving (4) in a distributed manner.

As it will be shown, an important feature of the proposed framework is that each EV adjusts its arrival rates only based on the updated prices that are dynamically announced by the station operators, and such distributed adjustment of arrival rates by the EVs leads to an optimal global welfare for the entire network.

By recalling that utility function $U^n$ is increasing in the arrival rates $\{\lambda_j\}$, decreasing in the LoLP $\{\beta_j\}$, concave in $\{\lambda_j\}$, and continuously differentiable in all of its arguments, the aggregate social welfare $R = \sum_{n=1}^{N} U^n$ is maximized when

$$\frac{\partial R}{\partial \lambda_j} = \frac{\partial U^n}{\partial \lambda_j} + \sum_{l=1}^{J} \sum_{s=1}^{J} \frac{\partial U^n}{\partial \beta_s} \cdot \frac{\partial \beta_s}{\partial \lambda_j} = 0. \quad (7)$$

Since the LoLP of each charge type depends only on the sum of the arrival rates of that type $\lambda_j = \sum_{n=1}^{N} \lambda^n_j$, from (7), we immediately have

$$\frac{\partial R}{\partial \lambda^n_j} = \frac{\partial U^n}{\partial \lambda^n_j} + \sum_{l=1}^{J} \sum_{s=1}^{J} \frac{\partial U^n}{\partial \beta_s} \cdot \frac{\partial \beta_s}{\partial \lambda^n_j} = 0. \quad (8)$$

Solving (8) yields a globally optimal set of arrival rates for the customers.

B. Local Problems

We show that the global optimal solutions yielded by (8) can be found in a distributed way, in which each charging station forms and solves a local problem. In order to lay the foundation, we first formulate the local problems in this section, and relegate establishing the connection between the local and global solutions to Section III-C. In order to enforce such arrival rates, the charging stations compute and announce prices for each customer type, which in turn forces the customers to operate at the desired (optimal) arrival rates. In order to solve this global problem and formalize such dynamic between posted prices and adjusted arrival rates, by recalling that $p_j$ is the price charged to customers of type $j$, we first define the following local optimization problem for each customer $n \in \{1, \ldots, N\}$ and service type $j \in \{1, \ldots, J\}$:

$$\tilde{U}^n(\lambda^n; \beta(\lambda)) \triangleq \frac{U^n(\lambda^n; \beta(\lambda))}{\text{gain}} - \sum_{l=1}^{J} \lambda^n_l (1 - \beta_l). \quad (9)$$

By solving this local optimization problem, customer $n$ computes its locally optimal and relevant arrival rates $\lambda^n = [\lambda^n_1, \ldots, \lambda^n_J]$. Specifically, the locally optimal arrival rates that maximize the local optimization problems $\{\tilde{U}^n\}_n$ for all $\forall n \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, J\}$ satisfy

$$\frac{\partial \tilde{U}^n}{\partial \lambda^n_j} - p_j (1 - \beta_j) = 0. \quad (10)$$

By solving (10), each customer $n$ can compute its own arrival rates based on the announced prices. The structure of the optimal local arrival rates depends on the selection of the utility function $U^n$, which is discussed in more detail in Section VI.

C. Connection Between Local and Global Solutions

Before proceeding to the details of the relevance of the globally optimal solution to these local solutions, it is noteworthy that individual customers do not have knowledge about the derivatives of their arrival rates with respect to LoLP $\beta_j$, which leads to local optimization problem [27]. Also, the size of the system is relatively large when compared to the demand of each individual customer, hence each EV does not have significant impact on the LoLP. Hence, by comparing (8) and (10), we can deduce that by the optimal prices for $j \in \{1, \ldots, J\}$ satisfy

$$p^*_j = -(1 - \beta_j)^{-1} \sum_{l=1}^{J} \sum_{s=1}^{J} \frac{\partial U^n}{\partial \beta_s} \cdot \frac{\partial \beta_s}{\partial \lambda_j} = 0. \quad (11)$$

When a customer of type $j$ is presented with price $p^*_j$, the solution of her local optimization problem (9) satisfies the conditions given in (7), which in turn guarantees that the solution of the global optimization problem in (8) is equivalent to the local ones given in (9) as established in [27]. Hence, the local optimal solutions become the equilibrium since no single customer can obtain a higher utility by deviating from its locally optimal solution. It is noteworthy that the prices $\{p^*_1, \ldots, p^*_J\}$ can be considered as congestion prices since each customer pays other customers for one unit of marginal decrease in
their utility due to the rise in the LoLP due to increase in the arrival rates.

IV. RESOURCE PROVISIONING FRAMEWORK

In the resource provisioning problem characterized in (6), the optimal solution $C^*$ increases monotonically as the target outage upper bounds increase. Hence, one naive approach to solve this problem is a linear brute force search, which can be computationally prohibitive for large-scale system. Instead, we aim to provide an established analytical connection between the optimal solution and the target outage requirements. This proposed approach, moreover, is versatile enough to be applied to other relevant optimization scenarios, e.g., integration of renewable generation or energy storage system and for varying customer demand.

In order to furnish the tools, we model the request arrivals of the customers as a collection of $J$ independent queues [28]. In this model, we start by analyzing the assumption that the grid resources are infinite and treat the case of finite resources in the next stage. Let $Q^\infty_j$ denote the number of customers of type $j$ requesting $b_j$ units of power concurrently. Further, let $S$ be the sum of offered load by the system that is given by

$$S = \sum_{j=1}^{J} b_j Q^\infty_j. \quad (12)$$

Due to Poisson model for the arrival rates, for the mean and the variance of $Q^\infty_j$, we have $\mathbb{E}[Q^\infty_j] = \text{var}[Q^\infty_j] = q_j = (\lambda_j / \mu_j)$. Therefore

$$\mathbb{E}[S] = \sum_{j=1}^{J} b_j q_j, \quad \text{and} \quad \text{var}[S] = \sum_{j=1}^{J} b_j^2 q_j. \quad (13)$$

In the case of finite resources $C$, we can reformulate the carried load on the system by defining $Q^j_C$ as the number of simultaneous customers requesting $b_j$ units of grid resources. Then, a customer of type $j$ would be blocked (loss-of-load event), hence do not receive service, if the load on the system, represented by $\sum_{j=1}^{J} Q^j_C$, is greater than available resources $C - b_j$. For a given set of charging rates $b = [b_1, \ldots, b_J]$ and average numbers of arrival rates $q = [q_1, \ldots, q_J]$, let us denote the LoLP by of customer type $j$ by $\beta_j : \mathbb{R}^J \rightarrow [0, 1]$, which captures the connection between the block probability on one hand, and $b$ and $q$, on the other hand.

By noting that the random number of customers for each class $Q^\infty_j$ are mutually independent Poisson random variables, we have

$$\beta_j(q, b) = \mathbb{P}\left\{ C - b_j < \sum_{j=1}^{J} b_j Q^j_C \right\} = \mathbb{P}\left\{ C - b_j < \sum_{j=1}^{J} b_j Q^\infty_j \leq C \right\}$$

$$\sum_{j=1}^{J} b_j Q^\infty_j \leq C \right\}$$

$$\mathbb{P}\left\{ C - b_j < \sum_{j=1}^{J} b_j Q^\infty_j \leq C \right\} \mathbb{P}\left\{ \sum_{j=1}^{J} b_j Q^\infty_j \leq C \right\}. \quad (15)$$

In the resource provisioning problem of interest, in order to meet multicloud QoS targets $\{\delta_j\}$, $C^*$ should be at least as large as the mean offered load on the system, that is $S = \sum_{j=1}^{J} b_j q_j$. However, customers arrive in a stochastic fashion, hence it is required to add extra capacity to accommodate the fluctuations beyond the average offered load. For this purpose, $C^*$ is set equal to the mean of the total load $\mathbb{E}[S]$ adjusted by an extra term which is a multiple of its variance denoted by $x \cdot \text{var}[S]$.

In this formulation, more stringent QoS targets lead to larger values of $x$. The objective of the resource provisioning problem is to provide a closed-form solution to (6). To this end, we first scale the system with $\zeta > 0$, according to which we have the following capacity for the network:

$$\tilde{C}(\zeta, x) = \zeta \sum_{j=1}^{J} b_j q_j + x \sqrt{\zeta \sum_{j=1}^{J} b_j^2 q_j} \quad (17)$$

and we have the following limiting result [22], [29]:

$$\lim_{\zeta \to \infty} \sqrt{\zeta} \beta_j(q, b) = \frac{b_j}{\sqrt{\sum_{j=1}^{J} b_j^2 q_j}} \phi(\psi(x)) \quad (18)$$

where $\phi(x) \triangleq \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $\psi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$. Now, let us define function $\psi$ as the inverse of $\phi / \psi$ for $\forall x$, that is

$$\frac{\phi(\psi(x))}{\psi(\psi(x))} = x. \quad (19)$$

Note that $\psi(\cdot)$ is a strictly decreasing function with $\psi(y) + y > 0, \forall y$ ([22], [29], and references therein). Then, the asymptotic behavior of the QoS constraint $\beta^*(q, b) \leq \delta_j$ results

$$x \geq \psi\left( \frac{\delta_j}{b_j} \right) \left[ \frac{\sum_{j=1}^{J} b_j^2 q_j}{b_j} \right]. \quad (20)$$

The provisioned grid resources should satisfy QoS targets for all classes. Hence, the inequality in (20) yields

$$x \geq \psi\left( \min_{1 \leq j \leq J} \frac{\delta_j}{b_j} \frac{\sum_{j=1}^{J} b_j^2 q_j}{b_j} \right). \quad (21)$$

By operating at the lower possible values of $x$, the minimum required amount of resources solving the provisioning problem in (6) is

$$C^* = \sum_{j=1}^{J} b_j \frac{\lambda_j}{\mu_j} + \psi\left( \min_{1 \leq j \leq J} \frac{\delta_j}{b_j} \frac{\sum_{j=1}^{J} b_j^2 \lambda_j}{\sum_{j=1}^{J} b_j^2 \lambda_j} \right) \left[ \frac{\sum_{j=1}^{J} b_j^2 \lambda_j}{b_j} \right] \quad (22)$$

where $\psi(\cdot)$ can be easily computed numerically by solving [30]

$$x^{-1} e^{-0.5\psi(x)^2} - \sqrt{2\pi} \text{erf}\left( \frac{1}{\sqrt{2}} \psi(x) \right) - x \sqrt{0.5\pi} = 0 \quad (23)$$

and it can be plugged back into (22) to compute the required capacity.

It is noteworthy that providing enough resources to meet the QoS targets of the most dominant classes, that are the ones with minimum $\delta_j/b_j$, is sufficient for the remaining customer classes.
V. COMPUTING LOLPs

Both the pricing-based control problem and the resource provisioning problem fall into multidimensional loss systems (or multirate Erlang loss systems), which are used to evaluate the guaranteed performance of networks with limited resources. In such systems, an arriving customer requesting for a certain amount of grid resources is either admitted to the system or is blocked and we are interested in computing LoLP as a function of system parameters.

Computing the LoLP function requires the analysis of the \( J \) independent time reversible Markov chains in which the state of the system is defined as the number of customers of each type, that is \( Q = \{ Q_1, \ldots, Q_J \} \) and the state space is denoted by \( \Omega \triangleq \{ Q : \sum_{j=1}^{J} b_j Q_j \leq C \} \). Let \( Q_j^C \) denote the maximum number of customers of type \( j \) that can be served simultaneously. Assuming the order \( b_1 \geq \cdots \geq b_J \geq 0 \), without loss of generality, provides that \( 0 \leq Q_j^C \leq \cdots \leq Q_1^C \). Then, the probability of being at state \( Q \) is [28]

\[
\pi(Q) = \prod_{j=1}^{J} \frac{q_j^{Q_j^C}}{Q_j^C!} e^{-\eta_j}.
\] (24)

Next, similar to (15), we condition on the finite capacity and compute a generic state \( Q \) probability distribution as

\[
\pi(Q) = \frac{\pi(Q)}{\sum_{Q \in \Omega} \pi(Q)}.
\] (25)

Next, let us define the blocking states for customer type \( j \) as

\[
\Psi_j = \left\{ Q : C - b_j < \sum_{k=1}^{j} b_k Q_k^C \leq C \right\}.
\]

Hence, (16) can be rewritten as

\[
\beta_j(q, b) = \sum_{s \in \Psi_j} \pi(s) = 1 - \sum_{s \notin \Psi_j} \pi(s)
\] (26)

where the second term represents the probability that the charging station has a total capacity of \( C - b_j \) instead of \( C \) and \( \pi(s) \) is the steady-state probability mass function. Furthermore, let us define function \( H(C, J) \) as

\[
H(C, J) \triangleq \sum_{Q \in \Omega} \prod_{j=1}^{J} \frac{q_j^{Q_j}}{Q_j!}
\] (27)

based on which one can compute the LoLP for class \( j \) explicitly as

\[
\beta_j(q, b) = 1 - \frac{H(C - b_j, J)}{H(C, J)}.
\] (28)

The set \( \{ Q : bQ \leq C \} \) in (27) contains all the states corresponding to which no outage occurs. While (28) provides an explicit representation for \( \beta_j \), the associated computation can be costly when the system capacity \( C \) is large. Considering a real-world scenario in which the number of classes \( J \) is typically varies between 3 and 5, and \( C \) is in the order of megawatts, computing LoLP could be easily carried out via the Kaufman–Roberts algorithm (Algorithm 1) which involves a simple recursion which computes the occupied resources [23]. Let \( c \) denote the amount of resources being in use and

\[
\alpha(c) \triangleq \mathbb{P}\{c \text{ units of power in use}\}.
\]

We subsequently have

\[
\alpha(c) = \sum_{Q \leq c} \frac{q_j^{Q_j}}{Q_j!} \cdot \frac{1}{H(C, J)}
\] (29)

and the LoLP corresponding to customer type \( j \) can be calculated by using

\[
\beta_j(q, b) = \sum_{i=0}^{b_j - 1} \alpha(C - i).
\] (30)

Note that the above derivations are based on the assumption that grid resources are discretized (e.g., 1 kW is considered as 1000 discrete serving units) and the interpretation of Algorithm 1 is that LoLP of customer type \( j \) equals to some of occupied states that are in the range of \( C - (b_j - 1), \ldots, C - 1, C \). Another important measure of interest for solving the optimal pricing problem given in (11) is the set of derivatives of LoLP with respect to offer load by each customer class. In the proposed multiclass customer model, there is no explicit formulas for performance measures (LoLP) in terms of input parameters (\( C, \lambda_j, \mu_j \), etc.). In this paper, we follow the methods based on convolution algorithms [31] to compute the derivatives of the LoLP. To that end, the derivative of the LoLP associated with customer type \( j \) with respect to the traffic intensity of another class \( j_1 \neq j_2 \) can be computed by using the function \( \alpha(\cdot) \) as follows:

\[
\frac{\partial \beta_j}{\partial q_{j_2}} = \alpha(C - b_{j_1}) + \alpha(C - b_{j_2} - 1) + \cdots + \alpha(C - b_{j_2} - b_{j_1} - 1) - (1 - \beta^{j_2}) \beta^{j_1}.
\] (31)

A useful property of the LoLP is the elasticity property, which entails that the sensitivity of customer LoLP \( j \) to customer class \( k \) is the same as the sensitivity of customer LoLP \( k \) to type \( j \) [32]. This property is given as \( \frac{\partial \beta_j}{\partial q_{j_2}} = \frac{\partial \beta_j}{\partial q_{j_1}} \). Notice that Kaufman–Roberts algorithm is used in both of the proposed frameworks and the complexity of this algorithm is \( O(CJ) \) which provides a considerable improvement over solving the LoLP through (28) which has a complexity of \( O(C^J) \) [23], [24]. For the pricing-based control

---

**Algorithm 1: Kaufman–Roberts Algorithm** [23]

Set \( \kappa(0) = 0 \) and \( \kappa(i) = 0 \) for \( i \in \mathbb{R}^- \)

for \( i=1 \) to \( C \) do

\( \kappa(i) = \frac{1}{i} \sum_{j=1}^{i} b_j \eta_{j}(i - b_j) \)

end for

Compute \( H = \sum_{i=1}^{C} \kappa(i) \)

for \( i=0 \) to \( J \) do

\( \alpha(i) = \frac{\kappa(i)}{\Pi \kappa} \)

end for

for \( j=1 \) to \( J \) do

\( \beta_j(q, b) = \sum_{i=C - b_j + 1}^{b_j - 1} \alpha(i) \)

end for
Fig. 6. Numerical evaluation-I: system capacity $C(t) = 450 + 50 \sin(2 \pi \tilde{k}/80)$, $\tilde{k} \in (0, 10, \ldots, 80)$ and $C$ is assumed constant for a duration of $T = 10$.
(a) Social welfare computation, (b) and (c) Optimal arrival rates ($\lambda_{1,2}$) with two classes.

Fig. 7. Numerical evaluation-II. (a) LoLP performance, (b) Social welfare computation, (c) Capacity planning for varying QoS targets with two classes.

problem, we compute the derivatives of multirate blocking probabilities using the convolution-based algorithm [31] and similar to Kaufman–Roberts algorithm, the order of complexity is $O(J)$.

VI. SIMULATION RESULTS

A. Case Study-I: Pricing-Based Control

In this section, we present a case study to evaluate the social welfare maximization problem of the pricing-based control framework presented in Section III. The parameter setting for our case study is as follows. For the ease of representation, we assume that there are two customer types, namely, fast charging customers (type-I) and slow charging customers (type-II). In order to mimic the typical charging rate for a fast dc charging (that is 50 kW), we set $b_1 = 50$ units. Also, charging duration takes around 20 min, so mean service rate is set as $\mu_1 = 3$. In a similar manner, we tune the parameters for the slow charging customers as $b_2 = 7$ units and $\mu_2 = 0.42$. For the utility functions, we adopt the widely-used logarithmic utility function [21], [26]. The utility function of a single customer increases with the arrival rate and decreases with the customer LoLP according to

$$U_j = \begin{cases} \sum_{j=1}^{J} \omega_j \log(1 + \lambda_j) - \theta_j \log\left(1 + \beta_j^\lambda C\right) \lambda_j > 0 \\ 0 \end{cases} \quad \lambda_j \leq 0$$

where $\omega_j$ and $\theta_j$ are the weights of each class. Note that the weight is higher for customer types with higher demand $b_j$, mainly because fast charging customers demand more resources and hence gains more utility. A simple example is presented to clarify the matters. Assume that system capacity $C = 500$ and the weights are chose as $\omega_1 = 20$, $\omega_2 = 10$, and $\theta_1 = 60$, $\theta_2 = 20$. Then, the social welfare could be maximized by setting the arrival rates to $\lambda_1 = 8.6638$ and $\lambda_2 = 5.2001$. For this arrival rate, prices $p_1$ and $p_2$ are computed to be 0.3197 and 0.2211 and the resulting LoLPs are $\beta_1 = 0.0097$ and $\beta_2 = 0.0009$, and the maximum utility is computed to be 59.1238 units. Notice that $\omega_j$ is chosen greater than $\theta_j$ so that station operator is motivated to provide a good level of QoS. We further explore the relationship between the optimal arrival rates and the social welfare. As shown in Fig. 6(a), if the arrival rates deviate from its optimal value, the social welfare reduces.

We proceed to provide more numerical evaluations. In the first setting, we present the relationship between system capacity and optimal arrival rates for time-dependent case, where station capacity varies over time and follows $C(k) = 450 + 50 \sin(2 \pi k/80)$ (due to grid conditions) and also $C(k)$ is assumed to be constant for every $T = 10$ duration. Results depicted in Fig. 6(b) shows that, it is more beneficial to accept fast charging customers as they improve the social welfare function more than slow charging ones. For the given set of arrival rates, the corresponding prices in (11) and LoLPs
are given in Figs. 6(c) and 7(a), respectively. Obviously, since fast charging uses more resources, the corresponding prices are higher and due to higher arrival rate the corresponding LoLP is higher than the slow charging customers. Finally, we present the corresponding social welfare in Fig. 7(b).

B. Case Study-II: Capacity Planning

We proceed to compute minimum amount of grid resources to provide QoS guarantees for two customer classes with the same set of parameters ($\mu_1 = 3$, $\mu_2 = 0.42$, $b_1 = 50$, $b_2 = 7$) for a wide range of QoS targets ($0.001 \leq \delta_1 \leq 0.05$, $0.001 \leq \delta_2 \leq 0.05$) with fixed arrival rate $\lambda_{1,2} = 5$. The results depicted in Fig. 7(c) presents that, since type 1 is the dominant class in (most) regions where $\delta_1 / b_1 < \delta_2 / b_2$, providing resources for class-1 already satisfies the QoS targets for class 2. Next, we compute the required capacity for different range of arrival rates and fixed QoS targets $\delta_1 = \delta_2 = 0.03$. Results depicted in Fig. 8(a) can be used as a guideline to choose the required capacity for given arrival rates.

We proceed to investigate the percentage of reduction in station capacity ($C$) for a given LoLP targets. The motivation is that instead of providing zero percent LoLP, station operator can sacrifice to reject small amount of customers and reduce the stress on the grid. We use the same parameter setting as above and computed the required capacity respect to almost zero LoLP ($\delta_{1,2} = 10^{-6}$). As presented in Fig. 8(b), even one percent QoS targets leads to significant savings in station capacity. Our final evaluation is on evaluating the system performance for nonhomogenous arrival rates. Let us assume that arrival rate for class-II is constant and $\lambda_2(k) = 10$ and $\lambda_1 = 10 + 2 \sin(2\pi k/80)$. The QoS targets are set as $\delta_1 = 0.04$ and $\delta_2 = 0.01$. Then station operator can provision the system according to peak hour demand (computed as $C = 683$) and hence meet the QoS requirements at all times. The results are presented in Fig. 8(c).

VII. CONCLUSION

In this paper, we provided two important design problems for EV charging stations with multiple classes of customers. In the first one, given a capacity of a station and infinite amount of customer request, we have provided a framework to compute the optimal arrival rates such that the total social welfare is maximized. This is a typical case for charging stations located in large cities. On the other hand, in the second case we have considered charging stations located in small cities with the primary concern to calculate the minimum amount of grid resources such that each customer class is ensured with a certain level of QoS targets.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Teleke for the fruitful discussions. The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] I. Bayram, G. Michailidis, M. Devetsikiotis, and F. Granelli, “Electric power allocation in a network of fast charging stations,” IEEE J. Sel. Areas Commun., vol. 31, no. 7, pp. 1235–1246, Jul. 2013.
[2] K. Clement-Nyns, E. Haesen, and J. Driesen, “The impact of charging plug-in hybrid electric vehicles on a residential distribution grid,” IEEE Trans. Power Syst., vol. 25, no. 1, pp. 371–380, Feb. 2010.
[3] A. Shortt and M. O’Malley, “Quantifying the long-term impact of electric vehicles on the generation portfolio,” IEEE Trans. Smart Grid, vol. 5, no. 1, pp. 71–83, Jan. 2014.
[4] S. W. Hadley, “Impact of plug-in hybrid vehicles on the electric grid,” U.S. Dept. Energy, Oak Ridge Nat. Lab., Oak Ridge, TN, USA, Tech. Rep. ORNL/TM-2006/554, Oct. 2006.
[5] B. Sanzhong, D. Yu, and S. Lukic, “Optimum design of an EV/PHEV charging station with DC bus and storage system,” in Proc. IEEE Energy Convers. Congr. Expo., Atlanta, GA, USA, Sep. 2010, pp. 1178–1184.
[6] M. de Freige, M. Ross, G. Joos, and M. Dubois, “Power AMP: energy ratings optimization in a fast-charging station for PHEV batteries,” in Proc. IEEE Int. Elect. Mach. Drives Conf., Niagara Falls, ON, Canada, May 2011, pp. 486–489.
[7] S. Junseok, A. Toliyat, D. Turtle, and A. Kwasiński, “A rapid charging station with an ultracapacitor energy storage system for plug-in electrical vehicles,” in Proc. Int. Conf. Elect. Mach. Syst., Incheon, Korea, Oct. 2010, pp. 2003–2007.
[8] I. Bayram, G. Michailidis, M. Devetsikiotis, and B. Parkhkhed, “Strategies for competing energy storage technologies for in DC fast charging stations,” in Proc. IEEE Int. Conf. Smart Grid Commun., Tainan City, Taiwan, Oct. 2012, pp. 1–6.
[9] K. Turitsyn, N. Sinitsyn, S. Backhaus, and M. Chertkov, “Robust broadcast-communication control of electric vehicle charging,” in Proc. IEEE Int. Conf. Smart Grid Commun., Washington, DC, USA, Oct. 2010, pp. 203–207.
[10] B. Sungwoo and A. Kwasiński, “Spatial and temporal model of electric vehicle charging demand,” IEEE Trans. Smart Grid, vol. 3, no. 1, pp. 394–403, Mar. 2012.
[11] H. Liang, I. Sharma, W. Zhuang, and K. Bhattacharya, “Plug-in electric vehicle charging demand estimation based on queueing network analysis,” in Proc. IEEE Power Eng. Soc. Gen. Meeting, Washington, DC, USA, Feb. 2014, pp. 1–5.

[12] D. Callaway and I. Hiskens, “Achieving controllability of electric loads,” Proc. IEEE, vol. 99, no. 1, pp. 184–199, Jan. 2011.

[13] X. Xi and R. Stioshansi, “Using price-based signals to control plug-in electric vehicle fleet charging,” IEEE Trans. Smart Grid, vol. 5, no. 3, pp. 1451–1464, May 2014.

[14] N. Rahbari-Asr, M.-Y. Chow, Z. Yang, and J. Chen, “Network cooperative distributed pricing control system for large-scale optimal charging of PHEVs/PEVs,” in Proc. IEEE Ind. Electron. Soc., Vienna, Austria, Nov. 2013, pp. 6148–6153.

[15] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, W. Vong, and J. Jatskevich, “Optimal real-time pricing algorithm based on utility maximization for smart grid,” in Proc. IEEE Int. Conf. Smart Grid Commun., Washington, DC, USA, Oct. 2010, pp. 415–420.

[16] I. Bayram, G. Michailidis, I. Papapanagiotou, and M. Devetsikiotis, “Decentralized control of electric vehicles in a network of fast charging stations,” in Proc. IEEE Int. Glob. Commun. Conf., Atlanta, GA, USA, Dec. 2013, pp. 2785–2790.

[17] I. Bayram, G. Michailidis, and M. Devetsikiotis, “Electric power resource provisioning for large scale public EV charging facilities,” in Proc. IEEE Int. Conf. Smart Grid Commun., Vancouver, BC, Canada, Oct. 2013, pp. 133–138.

[18] I. Bayram, M. Abdallah, and K. Qaraqe, “Providing QoS guarantees to multiple classes of EVs under deterministic grid power,” in Proc. IEEE Int. Energy Conf., Dubrovnik, Croatia, May 2014, pp. 1403–1408.

[19] A. Von Meier, Electric Power Systems: A Conceptual Introduction. Hoboken, NJ, USA: Wiley, 2006.

[20] M. Falkner, M. Devetsikiotis, and I. Lambadaris, “An overview of pricing concepts for broadband IP networks,” IEEE Commun. Surveys Tuts., vol. 3, no. 2, pp. 2–13, Feb. 2000.

[21] Z. Fan, “Distributed charging of PHEVs in a smart grid,” in Proc. IEEE Int. Conf. Smart Grid Commun., Brussels, Belgium, Oct. 2011, pp. 255–260.

[22] R. C. Hampshire, W. A. Massey, D. Mitra, and Q. Wang, “Provisioning for bandwidth sharing and exchange,” in Telecommunications Network Design and Management. New York, NY, USA: Springer, 2003, pp. 207–225.

[23] J. Kaufman, “Blocking in a shared resource environment,” IEEE Trans. Commun., vol. 29, no. 10, pp. 1474–1481, Oct. 1981.

[24] V. Iversen, A. Nilsson, M. Perry, and A. Gersht, On Multi-Rate Erlang-B Computations. Amsterdam, The Netherlands: Elsevier, 1999, pp. 1051–1060.

[25] J. K. MacKie-Mason and H. R. Varian, “Pricing congestible network resources,” IEEE J. Sel. Areas Commun., vol. 13, no. 7, pp. 1141–1149, Sep. 2005.

[26] C. Courcoubetis and R. Weber, Pricing Communication Networks: Economics, Technology and Modelling (Wiley Interscience Series in Systems and Optimization). New York, NY, USA: Wiley, 2003.

[27] C. A. Courcoubetis and M. I. Reiman, “Pricing in a large single link loss system,” in Proc. 16th Int. Teletraf. Congr., Edinburgh, U.K., Jun. 1999, pp. 737–746.

[28] L. Kleinrock, Theory, Volume 1, Queueing Systems. Hoboken, NJ, USA: Wiley, 1975.

[29] D. Mitra and J. Morrison, “Erlang capacity and uniform approximations for shared unbuffered resources,” IEEE/ACM Trans. Netw., vol. 2, no. 6, pp. 558–570, Dec. 1994.

[30] W. Tian and H. Perros, “Analysis and provisioning of a circuit-switched link with variable-demand customers,” in Managing Traffic Performance in Converged Networks. Berlin, Germany: Springer-Verlag, 2007, pp. 890–900.

[31] V. B. Iversen and S. Stepanov, “Derivatives of blocking probabilities for multi-service loss systems and their applications,” in Next Generation Teletraffic and Wired/Wireless Advanced Networking. Berlin, Germany: Springer, Sep. 2007, pp. 260–268.

[32] R. R. Mazumdar, “Performance modeling, loss networks, and statistical multiplexing,” Synth. Lect. Commun. Netw., vol. 2, no. 1, pp. 1–151, 2009.