In proceedings of IEEE Smart Grid Comm., Nov 3-6, 2014, Venice, Italy.

Optimal Charging Strategies for Electrical Vehicles under Real Time Pricing

Mohammad M. Karbasioun, Ioannis Lambadaris, Gennady Shaikhet, Evangelos Kranakis
Carleton University, Ottawa, ON, Canada
E-mails: {mkarbasi@sce, Ioannis@sce, gennady@math, kranakis@scs}.carleton.ca

***Abstract*—We address the problem of serving plug-in hybrid electric vehicles (PHEVs) in a charging station using a local storage energy unit, with finite capacity. Our goal is to find a control policy minimizing the operational cost of the charging station. We assume that the price of the electrical power unit is determined by a Real Time Pricing scheme, in which, the price of the electrical power is continuously updated with respect to the state of the grid at each time instance. We first model the charging station as a continuous time Markov Decision Process with three control actions: the probability of blocking new arrivals, the rate of charging the battery and the proportion of cars being served by the battery. By using dynamic programming, we prove the existence of the optimal policy minimizing a discounted cost over an infinite time horizon. We also show that it is stationary and bang-bang, i.e. the admissible action set in the optimal policy assumes only the extreme values in the action set.**

I. Introduction

A. Motivation

Plug-in hybrid electric vehicles (PHEVs) have attracted the interest of both industry and consumers. It is expected that PHEVs will take a 10% market share of new car sales by the year 2015 and can pass 50% by the year 2050 [1]. This interest in PHEVs is mainly due to high gasoline prices and public concerns about green house emissions. In practice, the introduction of PHEVs remains challenging. For example by deploying only night-time charging, an average size PHEV needs a battery with a capacity of 40 miles per full charge; this necessitates the deployment of relatively big and expensive batteries for PHEVs. However, using charging stations will decrease this value to 13 miles per full charge [2], which will decrease the size of the battery and consequently reduce PHEV prices in the market.

Charging a large number of PHEVs at the same time, will push the peak demand in the electrical grid. It will increase the need for extra supplementary power supplies during peak hours and hence decrease the efficiency of the grid. To decrease the peak demand, different "Demand Side Management (DSM)" [3] techniques are deployed. Without necessarily decreasing the total energy consumption, DSM "shifts" some consumption demands from peak times to "off-peak" times to flatten the consumption profile of the system as much as possible and consequently enhances the utilization of the total power grid. DSM uses different schemes to influence the consumption pattern. Among them are some pricing schemes to encourage consumers to change their consumption behaviour in order to enhance the utility of the whole electrical grid. Time of Use (TOU) scheme, is one of the most popular of such pricing schemes. In TOU scheme, different tariffs for the electrical power unit are set in different times of a day. For example in the Double Tariff scheme, which is very popular in the electricity market, consumers pay more per power unit during day-time in weekdays than during night-times and weekends.

In order to have better connection between the electricity prices and real-time situation of the grid, "Real Time Pricing" (RTP) scheme has been introduced. In RTP scheme, the price of an electrical power unit is continuously updated with respect to the state of the grid at each time instance and sent to the consumers in real time practically without delay [4]. Using RTP scheme, a utility can set the prices based on the state of the power generation, consumption rate and grid congestion at each time instance. Unfortunately due to limited flexibility of consumers, they don't respond to energy prices as much and show only a minor shift in their consumptions [5]. Therefore in addition to these types of incentives, we need to develop more complex and sophisticated control policies than just working with prices.

B. Related Research and Goal

Related work to our problem appears in [1], [6] and [7]. In [1], authors suggest a model for a charging station using a local storage unit, while the price of electrical power unit is fixed. In their model, the policy of serving demands is fixed and independent of the states. The main goal in [1] is to determine the proper size of the storage unit, by comparing the simulation results of the charging station model for different size of the storage unit. The simulation results mainly focus on the average cost and the blocking probability of the station.

In [6], a discrete time Markov Decision Process (MDP) model for serving electrical demands is presented. In this model it is assumed that the electrical demands can be fulfilled either by using the electrical supply from the grid or by drawing power from a local storage unit. In this model the demands arrive to the system according to a Poisson process, with identical but independent exponentially distributed random service times. In this model demands may not be blocked, but they can be delayed. The cost function is assumed to be a fixed convex function of consumed electrical power over the time. The goal in [6] is finding a control policy which minimizes the infinite horizon average cost of the system. Eventually an MDP problem is solved asymptotically for large storage capacity values.

In [7], authors introduce a model of serving electrical demands using power from either the grid or a local storage unit. The total amount of demands and the total available

Research supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC) grants.

power from the grid is assumed to be bounded as well as the capacity of the storage unit. However, since the total electricity demand is assumed to be less than the available power from the grid, there are no blocked demands. The resulting model follows a discrete time MDP and the goal is to minimize the expected total discounted cost of serving demands.

In this paper, we will present a model for controlling a charging station facility using a local storage energy unit with finite capacity. We take into account the price fluctuations for electrical power by considering a Real Time Pricing scheme. Our goal is to determine an optimal charging policy resulting in the minimum operational cost. To do so, we first model the charging station as a continuous time Markov Decision Process with three control actions at each state. We prove the existence of a stationary and bang-bang optimal policy minimizing a discounted cost over an infinite time horizon and further extend our results and determine the optimal policy minimizing the infinite horizon average cost. The rest of the paper is organized as follows: In Section II, we will propose a new model for a charging station. In Section III after introducing the proper cost functions and the optimal total discounted cost, we will determine the optimal policy minimizing a discounted cost over an infinite time horizon. In Section IV, we will define the infinite horizon average cost and then we will show how to find the average cost optimal policy from a discounted cost optimal policy. Finally in Section V, the performance of the optimal policy is compared against other policies.

II. Charging Station Model

Our model is motivated by the research in [1]. We assume vehicles arriving to a charging station, each carrying a certain charging demand. The arrivals are assumed to be Poisson with rate λ; the demands are independent and identically distributed (i.i.d) exponential random variables with rate μ (To simplify the analysis, we assume that the electrical power is quantized by the number of cars being charged.). The Markovian/memoryless assumptions for the arrivals and demands distributions, although may not be entirely realistic, they will nevertheless enable us to model our problem as a Markov Decision Process (MDP); this would not be possible for general arrival and demand distributions. We believe this approach may provide useful intuition to similar practical problems.

The charging station can draw the electrical power either directly from the grid or from a *local storage* facility (a battery). At any time the grid is able to charge at most S cars. By *level* of the battery we mean the number of cars that can be simultaneously charged from the battery at a given time. The level fluctuates between 0, indicating that the battery is empty, and R (the capacity of the battery) - the battery is fully charged. Both S and R are fixed/given deterministic parameters of the model. Let $i = i(t)$ denote the number of cars in the charging station at time t and $j = j(t)$ be the level of the battery at time t. The station can not have more than $S+j$ cars being charged at any given time, which is regulated by *blocking* the newly arrived cars. This implies

$$0 \leq j \leq R \qquad \text{and} \qquad 0 \leq i \leq S+j \tag{1}$$

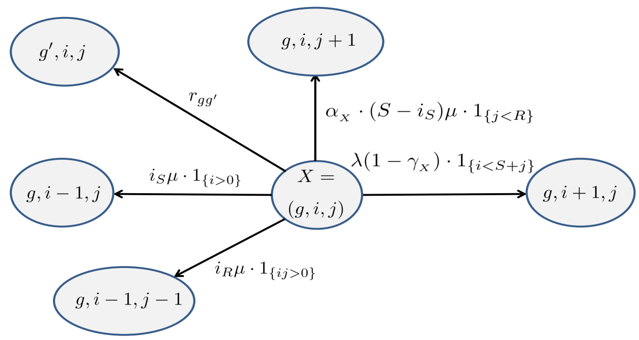

Fig. 1: The out-flow of the state g, i, j in the charging station

At any given time, a car can be served either by the grid or the battery. Letting i_S and i_R represent the number of cars being charged by the grid and battery respectively, we have

$$0 \leq i_S \leq S \quad \text{and} \quad 0 \leq i_R \leq j \quad \text{and} \quad i_R + i_S = i \tag{2}$$

If $i_S < S$ at some point of time, the remaining $S - i_S$ unused power units of the grid **could** be used to charge the battery, if it is not fully charged, i.e. $j < R$. The charging time of each level of battery when using a unit of electrical power is a random variable exponentially distributed with rate μ which is independent of charging times of other levels of battery and also independent of charging times of PHEVs. Therefore we assume that the charging times of each PHEV and each level of battery are i.i.d exponential random variables with the same rate μ. If the battery is being charged using $(S - i_S)$ units from the power grid, the charging rate of each level of battery is $(S - i_S)\mu$.

The state of the charging station at each time t, is determined by a three-dimensional vector $X = (g, i, j)$. Variables i and j are defined as above. An additional variable g denoting the price of each unit of electrical power in the grid, supports our assumption about RTP scheme for pricing. In particular, we assume that $g \in G$, where G, the power unit price set, is a finite discrete set composed of all possible values of g. Furthermore assume that the time the grid remains in each of the states from G is an exponential random variable with rate $r_g = \sum_{g' \neq g} r_{gg'}$, where $r_{gg'}$ is the rate of going from a state of price g to one of price g'. We assumed that new updated prices will be sent to the station using a high-speed connection without delay. Therefore at each time t, the state of the system, X, belongs to the state space

$$\mathcal{X} = \{(g, i, j) : g \in G,\ 0 \leq j \leq R,\ 0 \leq i \leq S+j\} \tag{3}$$

Now we are going to introduce controls to represent our model as a *controlled* Markov Process. At each time t, and state of the system $X = (g, i, j)$, we introduce a control vector $\mathbf{u}_{X,t} = (\gamma_{X,t}, i_{R_{X,t}}, \alpha_{X,t})$ in the following manner. The controls will be applied at transition instances so that an appropriate cost (to be introduced) is optimized.

The control variable $\gamma_{X,t} \in [0, 1]$ represents the fraction of arriving cars being blocked by the controller. It means upon arrival of a new car, the controller accepts this car with probability $(1 - \gamma_{X,t})$. Clearly, when $i = S + j$, $\gamma_{X,t} = 1$.

The second control variable $\alpha_{X,t} \in [0,1]$ will represent the proportion of available grid power to be used for charging the battery. So when there are $(S - i_S)$ power grid units available, the battery charge rate is equal to $\alpha_{X,t}(S - i_S)\mu$. Clearly, when $j = R$, $\alpha_{X,t} \equiv 0$. The last control variable is $i_{R_{X,t}}$. When entering a new state X, the controller splits the cars being charged in the station between the grid and the battery. In other words the controller determine i_R and i_S, independent of their values in previous states. Here we assume that there is no restriction to assign a car previously being charged by the grid, to be charged by the battery and vice versa. With respect to conditions in Equation (2), we have: $i_R \in \{(i-S)^+, ..., \min(i,j)\}$, where $(i-S)^+ = \max(0, i-S)$. Clearly, when $i = 0$ or $j = 0$, we have $i_{R_{X,t}} \equiv 0$ and $i_{S_{X,t}} \equiv i$. These definitions result in $\mathbf{u}_{X,t} = (\gamma_{X,t}, i_{R_{X,t}}, \alpha_{X,t}) \in \Omega$, where $\Omega = [0,1] \times \{(i-S)^+, ..., \min(i,j)\} \times [0,1]$. Figure 1 illustrates the transition diagram in the resulted Markov process. These rates form the generator matrix Q. The resulting controlled Markov process has a finite state-space.

III. Optimal Control Policy

We now need to introduce instantaneous cost function for our model. Given the state at time t is $X_t = (g,i,j)$ and the selected action vector is $\mathbf{u}_{X,t} = (\gamma_{X,t}, i_{R_{X,t}}, \alpha_{X,t})$. We define the operational grid cost incurred by the charging station during the time interval $[t, t+dt]$, as follows:

$$C_t^1(X, \mathbf{u}_{X,t})dt = (gi_{S_{X,t}} + g\alpha_{X,t}\ (S - i_{S_{X,t}}) - Vi)dt$$

where V is the revenue obtained by the charging station per unit time from each PHEV being charged at the station.

In addition, we are interested in taking into account the quality of service offered by the station, in terms of the number of blocked PHEVs. Let C_{bl} be the cost incurred per each blocked PHEV. Then the corresponding blocking cost in the time interval $[t, t+dt]$ is: $C_t^2(X, \mathbf{u}_{X,t})dt = C_{bl}\lambda\gamma_{X,t}dt$. Combining these two costs results in the following equation:

$$\begin{aligned} C_t(X, \mathbf{u}_{X,t})dt &= (C_t^1(X, \mathbf{u}_{X,t}) + C_t^2(X, \mathbf{u}_{X,t}))dt \\ &= (g \cdot i_{S_{X,t}} + g \cdot \alpha_{X,t} \cdot (S - i_{S_{X,t}}) \\ &\quad - V \cdot i + C_{bl}\lambda\gamma_{X,t})dt \end{aligned} \tag{4}$$

This cost function is stationary and independent of the time. So at each time t, we can write:

$$C_t(X, \mathbf{u}_{X,t})dt = C(X, \mathbf{u}_{X,t})dt \tag{5}$$

A control policy is a set of decision rules $\mathbf{z} = \{\mathbf{u}_{X,t} = (\gamma_{X,t}, i_{R_{X,t}}, \alpha_{X,t}) \in \Omega : \quad X \in \mathcal{X}, t\}$. Denote by $\mathcal{Z}$ the set of all policies. We are interested in finding a policy which minimizing the average incurred cost during the operating time of the system. In the sequel, we assume that the system operates over an infinite horizon time $[0, \infty)$.

Our problem can be modelled as a Markov Decision Process (MDP). Following standard practice in Markov Decision Processes, we will introduce a finite horizon δ-discounted cost problem and find its optimal control policy. Then the optimal policy of the infinite horizon problem can be derived as a limit of the optimal policies of a sequence of δ-discounted problems [8]. Let δ be the discount rate. Then given the initial state is X and the time horizon is $\{t : t > 0\}$, $\mathbf{J}_t^\delta(X)$, the minimum expected total δ-discounted cost is defined as follows:

$$\mathbf{J}_t^\delta(X) = \min_{\mathbf{z} \in \mathcal{Z}} \mathbf{E}_\mathbf{X}^\mathbf{z} \left(\int_0^t e^{-\delta\tau} C(X_\tau, \mathbf{u}_{X,\tau}) d\tau \right) \tag{6}$$

So the infinite horizon optimal δ-discounted expected cost is:

$$\mathbf{J}^\delta(X) = \min_{\mathbf{z} \in \mathcal{Z}} \mathbf{E}_\mathbf{X}^\mathbf{z} \left(\int_0^\infty e^{-\delta\tau} C(X_\tau, \mathbf{u}_{X,\tau}) d\tau \right) \tag{7}$$

If there exists a policy $\mathbf{z}^*$, resulting in $\mathbf{J}^\delta(X)$, this policy is called the optimal δ-discounted expected cost policy. This optimal policy satisfies the following optimality condition [9]:

$$\begin{aligned} \mathbf{J}_{t+dt}^\delta(X) &= \min_{\mathbf{u}_{X,0} \in \Omega} \{C(X, \mathbf{u}_{X,0})dt \\ &+ e^{-\delta dt} dt (\sum_{X' \in \mathcal{X}} \mathbf{P}(X_{dt} = X' | X_0 = X, \mathbf{u}_{X,0}) \mathbf{J}_t^\delta(X'))\} \end{aligned} \tag{8}$$

Lemma 1: For the cost function defined in (4), the optimal policy $\mathbf{z}^*$ with respect to equation (7) exists and it is stationary.

Proof: The state space $\mathcal{X}$ is finite and the resulting Markov chain (Figure 1) is irreducible. In addition, since we assumed that S and R are finite values, (Section II), the transition rates are finite and all of these transition rates are independent of time. Moreover from Equation (5), we know that the cost function $C(X, \mathbf{u}_{X,t})$ is independent of time. Therefore the optimal policy exists and it is stationary [8]. That is when at time t, the state of the system is $X_t = X$, the optimal action exists, is independent of t and only depends on the state of the system, X. So we have: $\mathbf{z}^* = \{\mathbf{u}_{X,t}^* = \mathbf{u}_X^* : \quad X \in \mathcal{X}, \quad t\}$. ∎

A. Deriving the Optimal Policy

So far it has been shown the optimal policy exists and it is stationary. Now in this subsection, we will show how to find the optimal policy. To do so first we relate the continuous time Markov chain $\{X_t : \quad t \geq 0\}$ to a suitable discrete time Markov chain. Then we will find the corresponding cost function and transition probabilities for the resulting discrete time Markov chain. Eventually we will apply them to the corresponding discrete time dynamic programming equations to find the optimal action for each state and hence the optimal policy. By using the method of "uniformization" [10], we can relate the continuous time Markov chain $\{X_t : \quad t \geq 0\}$ to a suitable discrete time chain $\{X_k : \quad k \geq 0\}$. Define the "total event rate" as follows: $\rho = \lambda + (S+R)\mu + \sum_{g \in G} r_g$. Let $0 = t_0 < t_1 < t_2 < \cdots < t_n < \cdots$ be the transition epochs with respect to possible transitions defined in the original Markov Chain shown in Figure 1. By suitably introducing "dummy" transitions, the inter-transition intervals become i.i.d exponential random variables with rate ρ. It can be shown [10] that for a policy $\mathbf{z} \in \mathcal{Z}$, the δ-discounted expected cost up to time t_n, i.e.: $\mathbf{E}_X^\mathbf{z} \left(\int_0^{t_n} e^{-\delta\tau} C(X_\tau, \mathbf{u}_{X,\tau}) d\tau \right)$ is equal to a cost:

$$\mathbf{E}_X^\mathbf{z} \left(\sum_{k=0}^{n-1} \beta^k C_d(X_k, \mathbf{u}_{X,k}) \right) \tag{9}$$

where $\{X_k : k \geq 0\}$ is the discrete time Markov chain obtained from $\{X_t : t \geq 0\}$ by using the uniformization technique, such that $X_k \triangleq X_{t_k}$. Equivalently $\mathbf{u}_{X,k} \triangleq \mathbf{u}_{X,t_k}$ is the action at k-th transition, where the discrete time Markov chain is at the state X. Furthermore in equation (9), $\beta \triangleq \dfrac{\rho}{\rho+\delta} < 1$. So $C_d(X_k, \mathbf{u}_{X,k})$, the discrete version of $C(X_\tau, \mathbf{u}_{X,\tau})$, is:

$$\begin{aligned} C_d(X_k, \mathbf{u}_{X,k}) = \;& g\ i_{S\,X,k} + g\ \alpha_{X,k}\ (S - i_{S\,X,k}) \\ -\;& V\ i_X + \lambda\ \gamma_{X,k} C_{bl} \end{aligned} \tag{10}$$

The expectation in (9) is taken with respect to the probability distribution associated to the discrete time Markov chain $\{X_k : \quad k \geq 0\}$ when the control policy $\mathbf{z}$ is used. Following are the resulting transition probabilities:

$$\begin{aligned} &\mathbb{P}^{\mathbf{z}}(X_{k+1} = (g', i', j') | X_k = (g, i, j), \mathbf{u}_{X,k} = (\gamma, i_R, \alpha)) \cdot \rho \\ &= \begin{cases} \lambda(1-\gamma) \cdot 1_{\{i<S+j\}}, & g' = g,\ i' = i+1, j' = j \\ r_{gg'}, & g' \neq g,\ i' = i, \quad\ j' = j \\ i_R \mu 1_{\{ij>0\}}, & g' = g,\ i' = i-1, j' = j-1 \\ i_S \mu 1_{\{i>0\}}, & g' = g,\ i' = i-1, j' = j \\ \alpha(S - i_S)\mu \cdot 1_{\{j<R\}}, & g' = g,\ i' = i+1, j' = j \\ \rho - \rho', & g' = g,\ i' = i, \quad\ j' = j \end{cases} \end{aligned} \tag{11}$$

Where $\rho' = \quad \lambda\ (1-\gamma) \cdot 1_{\{i<S+j\}} + \sum_{g' \neq g} r_{gg'} + i_R \mu 1_{\{ij>0\}} + i_S \mu 1_{\{i>0\}} + \alpha(S - i_S)\mu \cdot 1_{\{j<R\}}$.

For each initial state $X \in \mathcal{X}$, we can define the n-step optimal β-discounted expected cost as follows: $\mathbf{J}^\beta_n(X) = \min_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}^{\mathbf{z}}_X \left(\sum_{k=0}^{n-1} \beta^k C_d(X_k, \mathbf{u}_{X,k}) \right)$. Then the infinite horizon optimal β-discounted expected cost is defined as follows:

$$\mathbf{J}^\beta(X) = \min_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}^{\mathbf{z}}_X \left(\sum_{k=0}^{\infty} \beta^k C_d(X_k, \mathbf{u}_{X,k}) \right) \tag{12}$$

Since the state space $\mathcal{X}$ is finite, it can be shown [8] that $\lim_{n\to\infty} \mathbf{J}^\beta_n(X) = \mathbf{J}^\beta(X)$. Furthermore the optimal policy exists and it is stationary.

Theorem 1: In the optimal policy $\mathbf{z}^*$, the set of all admissible actions for $\mathbf{u}^*_X$ has the form $\mathcal{A} = \{0,1\} \times \{(i-S)^+, \min(i,j)\} \times \{0,1\}$. In other words the optimal policy actions attain the extreme values in their corresponding sets. Such policies are known as *bang-bang* policies.

Proof: To prove Theorem 1, first we will derive the dynamic programming equations for discrete Markov chain X and perform the necessary minimization with respect to γ, i_R and α. The optimal policy is obtained from the discrete time formulation of the dynamic programming equation [10], as follows:

$$\begin{aligned} \mathbf{J}^\beta_{k+1}(X) = \min_{\mathbf{u}_{X,0} \in \Omega} &\{ C_d(X, \mathbf{u}_{X,0}) \\ &+ \frac{\beta}{\rho} (\sum_{X' \in \mathcal{X}} \mathbf{P}(X_1 = X' | X_0 = X, \mathbf{u}_{X,0}) \mathbf{J}^\delta_k(X')\ ) \} \end{aligned} \tag{13}$$

Where $\mathbf{u}_{X,0}$ denotes the chosen action at $k = 0$, while the the system is in initial state X. However, since the optimal policy is stationary, for each $X \in \mathcal{X}$ we have: $\mathbf{u}_{X,0} = \mathbf{u}_X$. Now we can expand Equation (13), using $\mathbf{u}_X = (\gamma_X, i_{R\,X}, \alpha_X)$. For simplicity in our notations, in the following equations, we will replace $(\gamma_X, i_{R\,X}, \alpha_X)$ with (γ, i_R, α). Performing some easy algebra results in:

$$\begin{aligned} &\mathbf{J}^\beta_{k+1}(X) = \\ &\min_{\gamma \in [0,1]} \{ \lambda\gamma(C_{bl} - \frac{\beta}{\rho} 1_{\{i<S+j\}} (\mathbf{J}^\beta_k(g, i+1, j) - \mathbf{J}^\beta_k(g, i, j))\ ) \} \\ &+ \min_{i_R \in \{(i-S)^+, \cdots, \min(i,j)\},\ \alpha \in [0,1]} \{ \\ &i_S (g + \frac{\beta\mu}{\rho} 1_{\{i>0\}} (\mathbf{J}^\beta_k(g, i-1, j) - \mathbf{J}^\beta_k(g, i, j)) \\ &+ i_R \frac{\beta\mu}{\rho} 1_{\{ij>0\}} (\mathbf{J}^\beta_k(g, i-1, j-1) - \mathbf{J}^\beta_k(g, i, j)) \\ &+ \alpha(S - i_S)(g + \frac{\beta\mu}{\rho} 1_{\{j<R\}} (\mathbf{J}^\beta_k(g, i, j+1) - \mathbf{J}^\beta_k(g, i, j)))\} \\ &+ \textit{terms not depending on } (\gamma, i_R, \alpha) \end{aligned} \tag{14}$$

Equation (14) shows that the optimal choice for γ doesn't depend on α and i_R. We know that if $i = S + j$, then $\gamma \equiv 1$. So when $i < S + j$, the optimal choice for γ is obtained as follows:

$$\gamma = \begin{cases} 0, & (C_{bl}\ \rho - \beta(\mathbf{J}^\beta_k(g, i+1, j) - \mathbf{J}^\beta_k(g, i, j))\ \ ) \geq 0 \\ 1, & (C_{bl}\ \rho - \beta(\mathbf{J}^\beta_k(g, i+1, j) - \mathbf{J}^\beta_k(g, i, j))\ \ ) < 0 \end{cases} \tag{15}$$

To find the optimal choices for α and i_R, we should consider the second part of the Equation (14). First assume that $i \cdot j = 0$. Then as it is discussed in Section II, $i_R = 0$ and $i_S = i$. Also we know that if $j = R$, then $\alpha \equiv 0$. From Equation (2), we have: $(S - i_S) \geq 0$, independent of our choice for i_R. Therefore when $i \cdot j = 0$ and $j < R$, by using Equation (14), the optimal choice for α is as follows:

$$\alpha = \begin{cases} 0, & (g\rho + \beta\mu(\mathbf{J}^\beta_k(g, i, j+1) - \mathbf{J}^\beta_k(g, i, j))) \geq 0 \\ 1, & (g\rho + \beta\mu(\mathbf{J}^\beta_k(g, i, j+1) - \mathbf{J}^\beta_k(g, i, j))) < 0 \end{cases} \tag{16}$$

Now consider a case when $i > 0$ and $j = R$. Hence $\alpha \equiv 0$. Consequently by using Equation (14), the optimal choice for i_R is as follows:

$$i_R = \begin{cases} (i-S)^+, & -g\rho + \beta\mu\mathbf{J}^\beta_k(g, i-1, j-1) \\ & \qquad - \beta\mu\mathbf{J}^\beta_k(g, i-1, j) \geq 0 \\ & \\ min(i,j), & -g\rho + \beta\mu\mathbf{J}^\beta_k(g, i-1, j-1) \\ & \qquad - \beta\mu\mathbf{J}^\beta_k(g, i-1, j) < 0 \end{cases} \tag{17}$$

The most general case happens when $i \cdot j > 0$ and $j < R$. In this case Equation (14) becomes as follows:

$$\begin{aligned} &\mathbf{J}^\beta_{k+1}(X) = \min_{i_R, \alpha} \{ \\ &i_R(-g + \frac{\beta\mu}{\rho} (\mathbf{J}^\beta_k(g, i-1, j-1) - \mathbf{J}^\beta_k(g, i-1, j)) \\ &+ \alpha(S - i_S)(g + \frac{\beta\mu}{\rho} (\mathbf{J}^\beta_k(g, i, j+1) - \mathbf{J}^\beta_k(g, i, j)))\ \} \\ &+ \textit{terms not depending on } (i_R, \alpha) \end{aligned} \tag{18}$$

As we can see in Equation (18), only the second term depends on α. Now since $(S - i_S) \geq 0$ (Equation (2)), in order to minimize the RHS of Equation (18) with respect to α, we should minimize the following expression: $\alpha(S - i_S)(g + \frac{\beta\mu}{\rho}(\mathbf{J}_k^\beta(g,i,j+1) - \mathbf{J}_k^\beta(g,i,j)))$, which results in using the same criteria as Equation (16). Now if this criteria result in $\alpha = 0$, Equation (18) becomes as follows:

$$\begin{aligned}\mathbf{J}_{k+1}^\beta(X) &= \min_{i_R}\{\\ & i_R(-g + \frac{\beta\mu}{\rho}(\mathbf{J}_k^\beta(g,i-1,j-1) - \mathbf{J}_k^\beta(g,i-1,j));\}\\ &+ \textit{terms not depending on } (i_R)\end{aligned} \tag{19}$$

Equation (19) tells us that using the same criteria as Equation (17) gives us the optimal value of i_R, which minimizes Equation (18). On the other hand if using the criteria in Equation (16) results in $\alpha = 1$, then Equation (18) becomes as follows:

$$\begin{aligned}\mathbf{J}_{k+1}^\beta(X) =&\\ & \min_{i_R}\{i_R\frac{\beta\mu}{\rho}(\mathbf{J}_k^\beta(g,i-1,j-1) - \mathbf{J}_k^\beta(g,i-1,j)\\ &+ \quad \mathbf{J}_k^\beta(g,i,j+1) - \mathbf{J}_k^\beta(g,i,j)))\ \}\\ &+ \textit{terms not depending on } (i_R,\alpha)\end{aligned} \tag{20}$$

Therefore in cases when $i \cdot j > 0$ and $j < R$, if using criteria in Equation (16) results in $\alpha = 1$, to find the optimal choice for i_R, we should use the following criteria:

$$i_R = \begin{cases} (i-S)^+, & (\mathbf{J}_k^\beta(g,i-1,j-1) - \mathbf{J}_k^\beta(g,i-1,j) \\ & +\mathbf{J}_k^\beta(g,i,j+1) - \mathbf{J}_k^\beta(g,i,j))) \geq 0 \\ \\ min(i,j), & (\mathbf{J}_k^\beta(g,i-1,j-1) - \mathbf{J}_k^\beta(g,i-1,j) \\ & +\mathbf{J}_k^\beta(g,i,j+1) - \mathbf{J}_k^\beta(g,i,j))) < 0 \end{cases} \tag{21}$$

Equations (15), (16), (17) and (21) show that for every state $X \in \mathcal{X}$, there exists a set of admissible actions $\mathcal{A}$, as follows:

$$\mathcal{A} = \{0,1\} \times \{(i-S)^+, \min(i,j)\} \times \{0,1\} \tag{22}$$

such that $\mathbf{u}_X = (\gamma_X, i_{R_X}, \alpha_X) \in \mathcal{A}$. Denote by $\mathcal{Z}^\mathcal{A}$, the set of all admissible policies. Then for each $\mathbf{z}^\mathcal{A} \in \mathcal{Z}^\mathcal{A}$, including $\mathbf{z}^*$, we have: $\mathbf{z}^\mathcal{A} = \{\mathbf{u}_X \in \mathcal{A} : X \in \mathcal{X}\}$. Therefore instead of initial action set Ω (Section II), we are sure that the actions in the optimal policy belong to a much smaller set $\mathcal{A}$, which is indeed the boundaries of the initial set. In the literature, such policies are called *"bang-bang"* policies. So $\mathbf{z}^*$ is a bang-bang policy. ∎

Using dynamic equation (14) iteratively along with the criteria in optimal policy in Section III, we will find the infinite horizon optimal β-discounted expected cost $\mathbf{J}^\beta(X)$ and its corresponding optimal policy $\mathbf{z}^* \in \mathbf{Z}^\mathbf{A}$, with finite number of iterations [10].

B. Expected total β-discounted cost

$C_d(X, \mathbf{u}_X)$ (Equation (10)) is the cost of the discrete-time Markov chain $\{X_k : k \geq 0\}$, when it is in the state $X \in \mathcal{X}$ and action $\mathbf{u}_{X,k}$ is determined by stationary policy $\mathbf{z}$. Then the β-discounted expected cost of a stationary and time-invariant policy $\mathbf{z}$, when the initial state is $X \in \mathcal{X}$, is computed as follows:

$$\mathbf{J}_\mathbf{z}^\beta(X) = \mathbb{E}_X^\mathbf{z}\left(\sum_{k=0}^{\infty}\beta^k C_d(X_k, \mathbf{u}_{X,k})\right) \tag{23}$$

Define the column vectors $\mathbb{J}_\mathbf{z}^\beta$ and $\mathbb{C}_\mathbf{z}$ of size $|\mathcal{X}|$, where their elements are $\mathbf{J}_\mathbf{z}^\beta(X)$ and $C_d(X, \mathbf{u}_X)$ respectively. In [10](Page 43, Chapter 4), it is shown that for a stationary policy $\mathbf{z}$, the set of $|\mathcal{X}|$ linear equations:

$$(\mathbb{I}_{|\mathcal{X}|} - \beta\mathbb{P}^\mathbf{z})\mathbb{J}_\mathbf{z}^\beta = \mathbb{C}_\mathbf{z} \tag{24}$$

has a unique solution. It assures us that for a stationary policy $\mathbf{z}$ and for each state $X \in \mathcal{X}$, the Equation (23) will definitely converge to a unique value. So for each stationary and time-invariant policy $\mathbf{z}$, including the optimal policy $\mathbf{z}^*$, we can compute the expected total β-discounted cost as follows:

$$\mathbf{J}_\mathbf{z}^\beta(\mathbb{X}) = \mathbb{E}_\mathbb{X}^\mathbf{z}\mathbf{J}_\mathbf{z}^\beta(X) \tag{25}$$

The Expectation in Equation (25) is computed with respect to a given initial distribution $\mathbb{X}$. Therefore to compute the the expected total β-discounted cost of a stationary and time-invariant policy $\mathbf{z}$, instead of using Equation (23), we can use linear Equation (24) along with Equation (25). In Section V, we will compare the expected total β-discounted cost of the optimal policy $\mathbf{z}^*$ with that of certain other policies.

IV. Minimizing the infinite horizon Average Cost

For each stationary policy $\mathbf{z} = \{\mathbf{u}_X \in \Omega : X \in \mathcal{X}\}$, where we have $\mathbf{u}_{X,t} = \mathbf{u}_X$ for each $X \in \mathcal{X}$, its equivalent stationary discrete time Markov chain (Section III) is irreducible. Moreover since the state space $\mathcal{X}$ is finite (Section II), it is positive recurrent and hence ergodic [11]. Therefore the steady state probability distribution $\pi = \{\pi(X) : \quad X \in \mathcal{X}\}$ exists [11]-[10], and is the solution of $\pi = \pi\mathbf{P}$ subject to $\pi \cdot \mathbf{1} = 1$, where $\mathbf{P}$ is the transition probability matrix for the resulting discrete time Markov chain $\{X_k^\mathbf{z} : k \geq 0\}$ and $\mathbf{1}$ is a column vector of size $|\mathcal{X}|$ with all elements equal to 1. Similarly for the optimal policy $\mathbf{z}^* = \{\mathbf{u}_X^* \in \mathcal{A} : X \in \mathcal{X}\}$ in Section III, by finding each action vector $\mathbf{u}_X^*$ for each state $X \in \mathcal{X}$, the transition probability matrix $\mathbb{P}^{z^*}$ in (11) is fully determined.

For a policy $\mathbf{z}$, the average cost per unit time incurred by the discrete Markov chain $\{X_k^\mathbf{z} : k \geq 0\}$ is:

$$\mathbf{J}(\mathbf{z}) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}^\mathbf{z}\sum_{k=0}^{N-1} C_d(X_k, \mathbf{u}_{X,k}) \tag{26}$$

Considering the ergodicity of the Markov chain Equation (26) becomes as follows: $\mathbf{J}(\mathbf{z}) = \mathbb{E}^\mathbf{z} C_d(X, \mathbf{u}_X)$. In [10], it is shown that for continuous time Markov Decision Processes, $\{X_t : t > 0\}$, we have:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T C(X_t, \mathbf{u}_{X,t})dt = \mathbf{J}(\mathbf{z}) \tag{27}$$

Furthermore it is shown that as $\beta \to 1$, the optimal β-discounted expected cost policy $\mathbf{z}_\beta^*$, is then average cost optimal. Therefore the optimal average cost value is:

$$\mathbf{J}^* = \lim_{\beta\to 1}\mathbb{E}^{\mathbf{z}_\beta^*} C_d(X, \mathbf{u}_X) \tag{28}$$

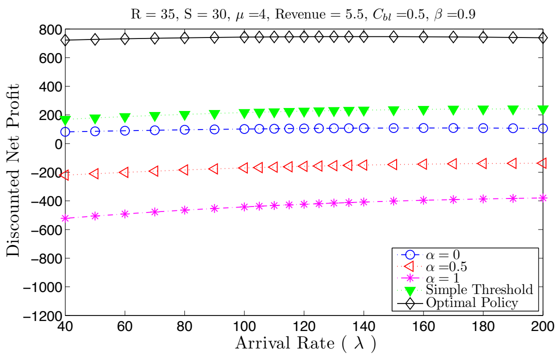

Fig. 2: The Discounted Utility $\mathbf{U}(z) = -\mathbf{J}(z)$ per unit time for optimal policy is compared to constant policies and threshold policy for different values of λ.

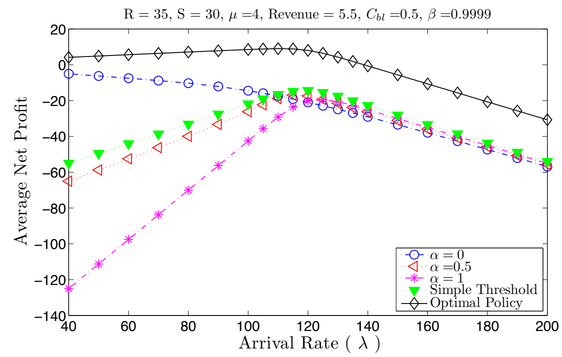

Fig. 3: The average Utility $\mathbf{U}(z) = -\mathbf{J}(z)$ per unit time for optimal policy is compared to constant policies and threshold policy for different values of λ.

The expectation in (28) is taken with respect to the stationary probability distribution associated to the discrete time Markov chain $\{X_k^* : k \geq 0\}$ obtained from the optimal policy $\mathbf{z}_\beta^*$ with transition probabilities obtained from Equation (11).

V. Simulation Results

In Figures 2 and 3, the output of the optimal policy in Section III, is compared to three α-Constant stationary policies for different values of λ. In each of these three policies, α is assumed to have the same value for all states $X \in \mathcal{X}$, i.e. $\mathbf{z}_\alpha = \{\alpha_X = \alpha, \quad \gamma_X = 1_{\{i=S+j\}}, \quad i_{R_X} = (i-S)^+ : X \in \mathcal{X}\}$ for $\alpha = 0, 0.5, 1$. For example $\mathbf{z}_\alpha$, for $\alpha = 1$ is the policy used in [1]. Furthermore it is compared to another policy, called *Simple Threshold*. In this policy for each state $X = (g, i, j) \in \mathcal{X}$, the action $\mathbf{u}_{X,t} = (\gamma_X, i_{R_X}, \alpha_X)$ is determined as follows: $\gamma_X = 1_{\{i=S+j\}}$, $\alpha_X = 1_{\{g \leq V\}}$ and $i_{R_X} = (i-S)^+ \cdot 1_{\{g \leq V\}} + \min(i,j) \cdot 1_{\{g>V\}}$, where V is defined in Section III and is the revenue obtained per unit time from each PHEV being charged at the station.

In these figures, the average Utility $\mathbf{U}$ per unit time is illustrated. $\mathbf{U}(z) = -\mathbf{J}(z)$ where $\mathbf{J}(z)$ is computed with respect to Equation (27) for each of these policies, including the optimal policy z^*. The parameters of the charging station are given as follows: $S = 30$, $R = 35$, $\lambda = 120$, $\mu = 4$, $V = 5.5$ and $C_{bl} = 0.5$. In addition, the given power unit price set is: $G = \{3, 5, 10\}$. All of theses unit prices are assumed to have the same mean times, e.g. 8 hours in a day, and equal transition probabilities to all other unit prices, i.e. $\frac{1}{2}$ in our setting. Therefore the transition rate between different unit prices is as follows: $r_{gg'} = \dfrac{1}{16}$ for all $g \neq g'$. Also in Figure 2, it is assumed $\beta = 0.9$. As we expected from Equations (12) and (28), the output of the optimal policy z^* dominates that of other policies.

In Figure 2, the expected total β-discounted cost (Section III-B) is considered. Note that since this cost depends on the initial state, in order to compare the output of different policies, we assume that the initial state is uniformly distributed in state space $\mathcal{X}$. Then after computing the β-discounted expected cost $\mathbf{J}_{\mathbf{z}}^\beta(X)$ (Equation (23)), we perform the averaging based on this assumption. On the other hand, in Figure 3, the long term average cost (Section IV) of the above policies are computed and compared with each other.

VI. Conclusion and Future Work

We considered the problem of optimally controlling a charging station of PHEVs in which PHEVs are charged either directly from the electrical grid or from a local storage unit. We modelled it as a continuous time Markov Decision Process in which the controller should decide on whether to accept or block the arriving demands, in addition to the charging rate of the local storage unit and the number of demands served by it. Then after introducing a proper cost function, we showed how to find the optimal policy. Finally we proved that the optimal policy is stationary and bang-bang.

As a possible future direction we will consider how to mathematically characterize the switching curve which separates different regions in the optimal policy. Furthermore as important future work, we will consider some different models for PHEV demands and also charging rates of the PHEVs and local storage unit.

References

[1] I. Bayram, G. Michailidis, M. Devetsikiotis, S. Bhattacharya, A. Chakrabortty, and F. Granelli, "Local energy storage sizing in plug-in hybrid electric vehicle charging stations under blocking probability constraints," in *IEEE SmartGridComm 2011 Track architectures and models*, 2011, pp. 78 –83.

[2] K. Morrow, D. Karner, and J. Francfort, "Plug-in hybrid electric vehicle charging infrastructure review," *US Departent of Energy-Vehicle Technologies Program*, 2008.

[3] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 3, pp. 381–388, 2011.

[4] S. Borenstein, "Customer risk from real-time retail electricity pricing: Bill volatility and hedgability," National Bureau of Economic Research, Tech. Rep., 2006.

[5] G. Barbose, C. Goldman, and B. Neenan. (2004) A survey of utility experience with real time pricing. Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory. Retrieved from: http://escholarship.org/uc/item/8685983c.

[6] I. Koutsopoulos, V. Hatzi, and L. Tassiulas, "Optimal energy storage control policies for the smart power grid," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011.

[7] P. Van de Ven, N. Hegde, L. Massoulie, and T. Salonidis, "Optimal control of residential energy storage under price fluctuations," in *ENERGY 2011 : The First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies*, 2011.

[8] S. Ross, *Introduction to Stochastic Dynamic Programming*. Academic Press, Inc. Orlando, FL, USA, 1983.

[9] Z. Rosberg, P. Varaiya, and J. Walrand, "Optimal control of service in tandem queues," *IEEE Transactions on Automatic Control*, vol. 27 , Issue: 3, pp. 600 – 610, 1982.

[10] P. Kumar and P. Varaiya, *Stochastic systems: estimation, identification and adaptive control*. Prentice-Hall, Inc., 1986.

[11] S. M. Ross, *Introduction to probability models*. Academic press, 2009.