Real-Time Vehicle-to-Grid Control Algorithm under Price Uncertainty

Wenbo Shi and Vincent W.S. Wong
Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver, Canada
E-mail: {wenbos, vincentw}@ece.ubc.ca

***Abstract*—The vehicle-to-grid (V2G) system enables energy flow from the electric vehicles (EVs) to the grid. The distributed power of the EVs can either be sold to the grid or be used to provide frequency regulation service when V2G is implemented. A V2G control algorithm is necessary to decide whether the EV should be charged, discharged, or provide frequency regulation service in each hour. The V2G control problem is further complicated by the price uncertainty, where the electricity price is determined dynamically every hour. In this paper, we study the real-time V2G control problem under price uncertainty. We model the electricity price as a Markov chain with unknown transition probabilities and formulate the problem as a Markov decision process (MDP). This model features implicit estimation of the impact of future electricity prices and current control operation on long-term profits. The Q-learning algorithm is then used to adapt the control operation to the hourly available price in order to maximize the profit for the EV owner during the whole parking time. We evaluate our proposed V2G control algorithm using both the simulated price and the actual price from PJM in 2010. Simulation results show that our proposed algorithm can work effectively in the real electricity market and it is able to increase the profit significantly compared with the conventional EV charging scheme.**

I. Introduction

Vehicle-to-grid (V2G) system enables the delivery of energy from the future electric vehicles (EVs) to the grid [1]–[3]. The batteries of the EVs in the V2G system can either provide power to the grid when parked or take power from the grid to charge the batteries. With V2G abilities, the EVs have a dual role in the electricity market. On one hand, they are power consumers when the batteries are being charged. On the other hand, they are power suppliers when they sell excessive energy from the batteries. As most of the vehicles are parked on an average of 96% of the time [4], they can be utilized as a power source besides transportation. However, each EV has a limited power capacity (10-20 kW) [2] and most of the services in the electricity market are carried out on a MW basis [5]. Therefore, an intermediate system called the V2G aggregator is needed to collect the small scale power from a large number of EVs in order to enter the electricity market.

The primary objective of the EVs when parked is to charge batteries before their next departure. The conventional approach would start charging at the maximum rate once plugged in until the expected state-of-charge (SOC) is reached [6]. With the development of V2G technology, bulk power selling and frequency regulation can be provided via the V2G aggregator in order to bring revenues for the EV drivers. The V2G aggregator controls the charging operations of the batteries of hundreds or thousands of EVs. As each EV is in different conditions (e.g., arrival and departure time, SOC, and capacity) from each other, an intelligent control algorithm is necessary for the aggregator to determine the control operation for the EVs in each hour (charging, discharging, or providing frequency regulation service) in a way that the profit of the EV is maximized subject to the EV's constraints (e.g., expected SOC must be reached).

There are a few V2G control algorithms proposed in the literature. The work in [7] considers the problem of maximizing the profits for the EV owners by selling excessive energy to the grid. Binary particle swarm optimization is used to determine if the EV should be charged, discharged, or in standby mode. Frequency regulation is integrated with the V2G system in [6]. The EVs in the V2G system can either be charged or provide frequency regulation. A dynamic programming (DP) algorithm is proposed to obtain the optimal control sequence for each EV. Both algorithms assume that the future electricity pricing information is given in advance based on a day-ahead pricing model.

With the recent development of smart grid technologies especially the communications infrastructure, real-time pricing [8]–[10] is becoming a promising scheme to increase power system reliability and efficiency by adjusting electricity prices according to the real-time supply and demand conditions [11]. The prices are usually high during peak hours and they change at different hours of the day to reflect the real-time cost of electricity. Real-time pricing has already been implemented in some places (e.g., Illinois Power Company in Chicago [8]). With real-time pricing, the electricity price can be determined by the utility company just a few minutes prior to the beginning of each hour. The customers can receive the hourly pricing information and respond to it by adjusting their electricity usage. However, real-time pricing creates great challenges for customers as they are facing not only the hourly decisions making problem but also the uncertainty of the future electricity prices.

The issue of price uncertainty brought up by real-time pricing has recently been studied in demand response in order to solve the decision making problem of whether customers should consume the energy now at current price or shift the demand to the future at unknown prices [12], [13]. In [12], a real-time demand response algorithm operating on a daily

24-hours horizon is proposed, where the price uncertainty is considered in the model via robust optimization. Another method to tackle the price uncertainty issue can be found in [13], where reinforcement learning is used in the residential demand response algorithm. Both [12] and [13] deal with the real-time demand response problem, while we consider the same price uncertainty problem in the context of V2G control.

In this paper, we consider the real-time V2G control problem under price uncertainty. We propose a novel V2G control algorithm that learns from past experiences and automatically adapts to the unknown pricing information and makes optimal hourly control decisions. The contributions of this paper are as follows:

- We model the V2G control problem as a Markov decision process (MDP), where the price uncertainty is taken into account by maximizing the long-term objective function. The decisions at current time are made with consideration of future profits.
- We then propose an online learning algorithm to automatically control the EVs in response to the hourly electricity prices. The proposed algorithm can adapt to the changing pricing information.
- We evaluate the proposed algorithm using both the simulated price and the actual price. Simulation results show that the proposed algorithm can increase the profit significantly and it is effective under real market conditions.

The real-time availability of the pricing information brought by the recent development of smart grid technologies differs our proposed V2G control algorithm from previous works [6], [7] which require the pricing information of the whole parking time in advance. Our proposed algorithm is based on real-time pricing, which is different from algorithms [6], [7] based on day-ahead pricing. Real-time pricing and day-ahead pricing are two different pricing models in the electricity market. Therefore, our proposed algorithm has a different application background than the previous algorithms. Modeling the time series of electricity price as a Markov chain is not new [14] and it can also be found in the real-time demand response algorithm [13], but the idea of applying it to the design of V2G control algorithm under price uncertainty is novel. In addition to the modeling of price uncertainty, our proposed algorithm is also different from the existing V2G control algorithms [6], [7] in that both frequency regulation and bulk power selling are considered.

The rest of the paper is organized as follows. Our system model is described in Section II. The MDP formulation of the V2G control algorithm under price uncertainty is discussed in Section III. The Q-learning algorithm to solve the MDP problem is presented in Section IV. Simulation results are provided in Section V. Conclusions are given in Section VI.

II. System Model

This section describes the real-time V2G control problem. It also introduces an overview of the V2G system and the objective of the system design, which will be used in the following MDP formulation.

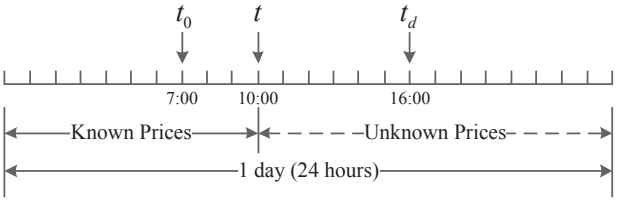


Fig. 1. The electricity price uncertainty in the V2G control problem.

A. System Overview

As we discussed in Section I, the small scale power capacity of a single EV's battery requires the V2G aggregator to gather the large scale power from hundreds or thousands of EVs in order to sell the V2G power or to provide the frequency regulation service in the electricity market. In the V2G system, all the control operations are initiated by the aggregator using the control algorithm. With real-time pricing program implemented, the V2G aggregator receives the pricing information a few minutes (e.g., 10 min) prior to the beginning of each hour. Using this pricing information, the V2G control algorithm is run for each EV, which will stay parked in the next hour, to find whether the aggregator should charge its battery for the next trip, discharge for selling excessive power, or use its available power capacity to provide frequency regulation service. After gathering the control operations of all the EVs, the V2G aggregator sends the contract information (i.e., total amount of power to buy, total amount of power to sell, and total capacity for frequency regulation) to the utility company (e.g., 5 min before the coming hour). The above communications process is carried out at the beginning of each hour. In the following hour, the contracted buying and selling power will be dispatched. An energy management system from the utility company will dispatch appropriate regulation signals to the V2G aggregator based on the contracted capacity using its own algorithm.

B. Objective

We now describe the V2G control problem under price uncertainty at time t. Assume that the EV arrives at time t_0 and will depart at time t_d as shown in Fig. 1. The time indices in our discussions are all assumed to be the beginning of the hour. The pricing information (i.e., the electricity price and the frequency regulation price) and the control decisions for the previous hours from t_0 to $t-1$ are known. The pricing information for the current hour is known (e.g., a few minutes prior to t). The prices for the following parking hours $(t, t_d-1]$ are unknown data and this uncertainty needs to be modeled. The control decision for time t needs to be determined by the control algorithm after it receives the pricing information for the current hour.

The objective of the V2G control algorithm is to maximize the profit for the EV owner which refers to the payment from the utility company by selling power and providing frequency regulation service minus the costs of purchasing power from the grid. To calculate the profit, we need to first understand how the payments are made. The payment from bulk power selling is equal to the product of the electricity price and

the amount of energy sold. Another source of payment is frequency regulation. Unlike bulk power selling, the payment from regulation is made according to the power capacity provided instead of the amount of real dispatched energy. In fact, the SOC of the battery will be affected due to the regulation signals. However, as the fluctuations of the positive and negative power deviations in regulations are uniformly distributed, the total amount of energy flows into the grid is equal to the amount of energy flows out of the grid in the long run [15]. Therefore, we assume that the SOC stays the same when providing frequency regulation service [6].

III. V2G Control Problem As An MDP

This section describes formulation of the V2G control problem under price uncertainty as an MDP. An MDP is completely described through its state space, action space, system dynamics, and value function. The following definitions are the foundation for the Q-learning algorithm to be discussed in the next section.

A. State Space

We consider the V2G control problem for a single EV arriving in the V2G aggregator at discrete time t_0. We assume that once the EV is parked, its departure time t_d and the expected SOC at departure B are known. The hourly electricity pricing signals are indicated by a vector $\mathbf{p}_t = [p_t^c \ p_t^r]$, where p_t^c is the market price for purchasing and selling electricity in $\$/\mathrm{kWh}$ and p_t^r is the price for providing frequency regulation in $\$/\mathrm{kWh}^{-1}$. Note that kWh^{-1} is the regulation power capacity contracted for an hour and should not be confused with kWh which is the energy unit. Let $0 \leq b_t \leq 1$ be the SOC which is defined as the percentage of the battery power capacity that are available at time t and l_t denote the time left for departure.

Let $s_t \in \mathcal{S}$ denote the state of the system at time t. We define $s_t = [\mathbf{p}_t \ b_t \ l_t] \in \mathcal{S}$, where $\mathcal{S}$ is the state space of the V2G control problem. $\mathcal{S}$ is the composite space comprising of the pricing space $\mathcal{P}$, the SOC space $\mathcal{B}$, and the remaining time space $\mathcal{L}$, i.e., $\mathcal{S} = \mathcal{P} \times \mathcal{B} \times \mathcal{L}$, where $\times$ denotes the Cartesian product. Note that there is a terminal state which ends the MDP when $t = t_d$.

B. Action Space

The action in the V2G control problem can be interpreted as choosing one control operation from the action space $\mathcal{A} = \{\text{charging, discharging, regulation}\}$. However, due to the constraints in the V2G control problem, not all the actions can be performed at a given state. Let $a_t \in \mathcal{A}^{s_t}$ denote the action, where $\mathcal{A}^{s_t}$ is the set of all possible actions given the state of the system.

$\mathcal{A}^{s_t}$ is limited by two types of constraints in the V2G control problem. The first constraint is that the EV must be charged to the expected SOC at departure. We assume that when the EV arrives, the driver will notify the expected SOC B and departure time t_d, which are essential information for the control algorithm, to the aggregator. This mandatory notification can be achieved by signing a contract by the EV driver to guarantee that the EV would be plugged in during the notified period of time in return of some incentives [16]. However, it still may happen that the EV leaves before the expected departure time. Under this circumstance, the battery of the EV may not be charged enough for the expected SOC even though it has been plugged in for a long time due to the discharging. Some V2G control algorithms restrict that once the expected SOC is reached, the EV cannot be discharged any more as way to deal with early departure [7]. We do not adopt this approach in our algorithm and leave the responsibility to the driver to make sure that the EV departs as expected.

Let C denote the charging rate which is the percentage of the battery energy capacity that can be charged per hour. We assume the discharging rate is the same as the charging rate but with an opposite direction which is denoted by $-C$. The action space satisfying the departure constraint can be expressed as

$$\mathcal{A}_1^{s_t} = \begin{cases} \{\text{charging, discharging, regulation}\}, \\ \qquad \text{if } l_t \geq \left\lceil \dfrac{B - b_t}{C} \right\rceil + 2, \\ \{\text{charging, regulation}\}, \\ \qquad \text{if } \left\lceil \dfrac{B - b_t}{C} \right\rceil < l_t < \left\lceil \dfrac{B - b_t}{C} \right\rceil + 2, \\ \{\text{charging}\}, \text{ if } l_t \leq \left\lceil \dfrac{B - b_t}{C} \right\rceil, \end{cases} \tag{1}$$

where $\lceil \cdot \rceil$ is the ceiling function.

The second constraint is the energy constraint. The V2G power flow from the EV to the grid is limited by not only the periphery circuits, but also the SOC. The battery management system would protect the battery by restricting charging and discharging when the SOC is approaching the maximum and minimum, respectively. Therefore, we forbid the battery from being charged and discharged when the SOC is approaching the maximum (e.g., 95%) and minimum (e.g., 5%), respectively. The action space satisfying the energy constraint can be obtained as

$$\mathcal{A}_2^{s_t} = \begin{cases} \{\text{discharging, regulation}\}, & \text{if } b_t \geq 95\% - C, \\ \{\text{charging, regulation}\}, & \text{if } b_t \leq 5\% + C. \end{cases} \tag{2}$$

The action space given the current state s_t satisfying both constraints would be the intersection of the sets $\mathcal{A}_1^{s_t}$ and $\mathcal{A}_2^{s_t}$

$$\mathcal{A}^{s_t} = \mathcal{A}_1^{s_t} \bigcap \mathcal{A}_2^{s_t}. \tag{3}$$

When the system is in state $s_t \in \mathcal{S}$, a finite number of possible actions which are elements of the set $\mathcal{A}^{s_t}$ can be taken.

At each time step, the control algorithm implements a randomized policy which is a discrete probability function of the state and the possible actions. The randomized policy at time t is denoted $\pi_t(s, a)$ which is the probability that action a is taken when the state is s at time t.

We define the function $\Psi(a_t), a_t \in \mathcal{A}^{s_t}$ which returns the

amount of energy dispatched from the grid by action a_t

$$\Psi(a_t) = \begin{cases} EC, & \text{if } a_t = \text{charging}, \\ 0, & \text{if } a_t = \text{regulation}, \\ -EC, & \text{if } a_t = \text{discharging}, \end{cases} \tag{4}$$

where E is the battery capacity of the EV in kWh.

C. System Dynamics

For a given policy, the evolution of an MDP is characterized by the transition probability

$$\mathbb{P}(s_{t+1} \mid s_t, a_t), \tag{5}$$

for some $s_t, s_{t+1} \in \mathcal{S}$, $a_t \in \mathcal{A}^{s_t}$, and $t = \{t_0, t_0+1, \ldots, t_d - 1\}$.

The evolutions of b_t and l_t are given by

$$\mathbb{P}(b_{t+1} \mid b_t) = \begin{cases} 1, & \text{if } b_{t+1} = b_t + \Psi(a_t), \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

and

$$\mathbb{P}(l_{t+1} \mid l_t) = \begin{cases} 1, & \text{if } l_{t+1} = l_t - 1, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

Therefore, we can obtain the system state transition probability as

$$\mathbb{P}(s_{t+1} \mid s_t, a_t) = \mathbb{P}(b_{t+1} \mid b_t)\mathbb{P}(l_{t+1} \mid l_t)\mathbb{P}(\mathbf{p}_{t+1} \mid \mathbf{p}_t). \tag{8}$$

Unfortunately, the underlying statistical structure of the pricing information is difficult to estimate and it is subject to the market conditions. Hence, $\mathbb{P}(s_{t+1} \mid s_t, a_t)$ is unknown and may change with time.

D. Value Function

Let finite reward $r(s_t, a_t)$ be the instantaneous revenue of taking action a_t at state s_t. We define the reward as the financial revenue for the EV owner. The reward is equal to the energy payment of charging and discharging or the capacity payment of providing frequency regulation. As the charging/discharging operations are performed at the rate C, the corresponding rewards are calculated based on the energy flow $\Psi(a_t)$. Note that the reward is the negative when the owner pays money and positive when the owner gains money. The payment of regulation is made using the maximum power capacity. To sum up, the reward function is defined as

$$r(s_t, a_t) = \begin{cases} -p_t^c EC, & \text{if } a_t = \text{charging}, \\ p_t^r EC, & \text{if } a_t = \text{regulation}, \\ p_t^c EC, & \text{if } a_t = \text{discharging}. \end{cases} \tag{9}$$

We adopt the expected total reward as the optimization criterion in the V2G control problem. Let the total reward conditioned on the initial state s_{t_0} under a given policy π be defined as

$$R_{s_{t_0}}^{\pi} = \mathbb{E}_{\pi}\left\{ \sum_{t=t_0}^{t_d-1} r(s_t, a_t) \mid s_{t_0} \right\}, \tag{10}$$

where the expectation is over randomized actions a_t and system state s_t evolution for $t = \{t_0, t_0+1, \ldots, t_d - 1\}$. The objective is to compute the optimal control policy π^* that maximizes the expected total reward (10).

IV. The Q-learning Algorithm

Solving (10) for the optimal control policy π^* generally requires the knowledge of the transition probability distribution of $\mathbf{p}_t$. Unfortunately, the underlying structure of the energy price is difficult to estimate and may change with market conditions. To address this challenge, we use an online learning algorithm. Among many such algorithms, we choose the Q-learning algorithm [17], [18] for its simplicity.

The Q-learning algorithm estimates the value of the optimal action-value function, denoted Q^*, and defined as

$$Q^*(s,a) = \max_{\pi} \mathbb{E}_{\pi}\left\{ \sum_{k=0}^{t_d-t-1} r(s_{t+k}, a_{t+k}) \mid s_t = s, a_t = a \right\}. \tag{11}$$

To approximate the optimal action-value function Q^*, the Q-learning algorithm learns from experiences and updates at each time step. Let a^* be the greedy action which maximizes $Q(s_t, a)$. At each time step, the Q-learning algorithm chooses a^* most of the time, but with probability ϵ it selects an action at random. The randomized policy $\pi_t(s_t, a), \forall\ a \in \mathcal{A}^{s_t}$ is defined by

$$\pi_t(s_t, a) = \begin{cases} 1 - \epsilon + \dfrac{\epsilon}{|\mathcal{A}^{s_t}|}, & \text{if } a = a^*, \\ \dfrac{\epsilon}{|\mathcal{A}^{s_t}|}, & \text{if } a \neq a^*, \end{cases} \tag{12}$$

where ϵ is a small number which ensures that all state-action pairs can be visited and updated.

The instant reward $r(s_t, a_t)$ can be obtained after a_t is executed and the system state transits to s_{t+1}. The learned Q function can be updated by

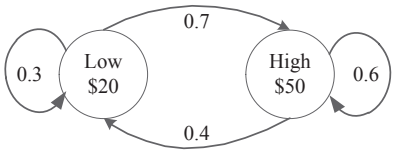
$$\begin{aligned} Q(s_t, a_t) := Q(s_t, a_t) + \alpha_t [r(s_t, a_t) \\ + \max_{a \in \mathcal{A}^{s_{t+1}}} Q(s_{t+1}, a) - Q(s_t, a_t)], \end{aligned} \tag{13}$$

where $0 < \alpha_t < 1$ is the learning rate. It determines how the new information is averaged with existing estimate. It can be shown that choosing the learning rate $\alpha_t = \frac{1}{t}$ under the usual stochastic approximation conditions and if all state-action pairs are continuously updated, Q_t converges to Q^* with probability 1. However, using this decreasing learning rate means that the algorithm will respond less to the environment as the time goes on, which is not helpful in the V2G control problem to capture the changing transition probabilities of the prices. Alternatively, we choose α_t to be a small constant number. It can be shown in [19] that Q_t converges to a region near the optimal Q^* using a constant learning rate. Although some estimation errors cannot be avoided using this method, it is able to adapt to the changes of the underlying statistical structure of the environment.

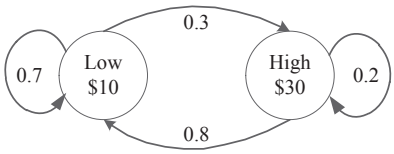
A complete description of the Q-learning algorithm for solving the real-time V2G control problem can be found in

Algorithm 1 - The Q-learning Algorithm: Executed by the V2G aggregator when the EV arrives.

1: $t := t_0$.
2: The EV informs t_d and B to the aggregator.
3: Initialization: $l_{t_0} = t_d - t_0$.
4: **Repeat**
5: Receive the real-time pricing information $\mathbf{p}_t$.
6: $a^* = \arg\max_a Q(s_t, a)$.
7: Choose a_t from policy π_t defined by (12).
8: Take action a_t and obtain the reward $r_t(s_t, a_t)$.
9: $b_{t+1} = b_t + \Psi(a_t)$, $l_{t+1} = l_t - 1$.
10: Update the Q function using (13).
11: $t := t + 1$.
12: **Until the EV departs**



(a) p_t^c



(b) p_t^r

Fig. 2. States and transition probabilities for (a) the market price p_t^c and (b) the regulation price p_t^r.

Algorithm 1. The algorithm is run for every EV in the V2G aggregator when it arrives. Upon the EV arriving, the driver needs to inform the expected departure time and expected departure SOC to the aggregator. At the beginning of the following hours, the algorithm will receive the real-time pricing information from the utility company (Step 5). Step 7 finds the control action to be taken at the current state based on the current estimate of the Q function. The aggregator executes the control action in Step 8 and obtains the reward. The system evolves after the action is implemented in Step 9. Step 10 updates the Q function. Steps 5-11 are repeated during the parking hours until the EV departs. Note that the Q function is a global variable stored in the V2G aggregator so that its value can be accessed and updated by all the EVs. The initial value of the Q function can be arbitrary.

V. Performance Evaluation

In this section, we present the simulation results for the proposed real-time V2G control algorithm. For simplicity, we consider the case where the proposed V2G control algorithm is run for a single EV on different days with exactly the same conditions. The battery energy capacity is $E = 20$ kWh. The charging rate $C = 0.1$. We assume that the EV arrives at 18:00 in the afternoon with arriving SOC of 40% every day and departs at 8:00 next morning with expected departure SOC of 70%. In the simulation, we consider two different scenarios.

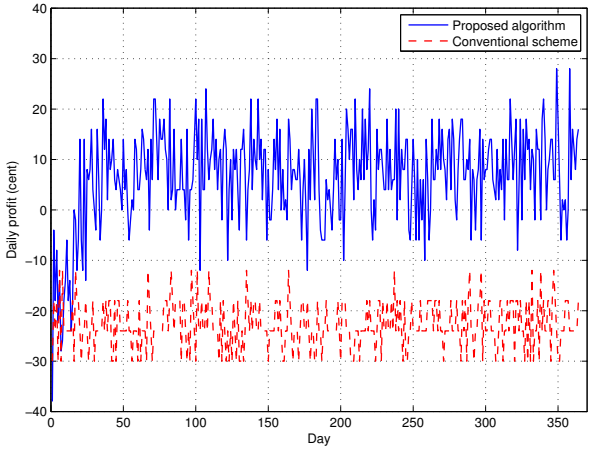


Fig. 3. The comparison between the daily profit of the proposed V2G control algorithm and the conventional charging scheme using the simulated price.

In the first scenario, we generate the electricity prices using a two-state Markov chain. In the second scenario, we use the actual real-time location marginal price (LMP) and regulation market clearing price (RMCP) from PJM Interconnection [20] which is a major regional transmission organization operating the wholesale electricity market in the US. The simulated prices and actual prices are all given for 24 hours a day on a MW scale. Note that the previously proposed V2G control algorithms [6], [7] cannot be directly compared with our proposed algorithm due to the different pricing models.

A. Simulated Price

We first implement our proposed V2G control algorithm using the electricity prices generated by a Markov chain. The Markov chain for p_t^c and p_t^r are shown in Fig. 2. The prices are chosen as the typical values from the actual price data from PJM. The initial state is randomly selected. The time period we run for the proposed algorithm is one year. The parameters in the algorithm are chosen as $\alpha_t = 0.05$ and $\epsilon = 0.02$ in the simulation.

The results of the daily profit of the proposed algorithm and the conventional scheme, which charges the battery at the maximum rate to the expected SOC once it is plugged in, are shown in Fig. 3. We can see the learning process of our proposed algorithm from the figure. The proposed algorithm does not show advantages over the conventional scheme at first. As it learns from the experiences and adapts to the changing prices, the daily profit of the proposed algorithm starts to increase. The proposed algorithm constantly outperforms the conventional scheme by a significant margin after the learning process which lasts about 25 days. If we compare the average daily profit after the learning process, the results would be $\$0.079$ and $-\$0.23$ for the proposed algorithm and the conventional scheme, respectively. The positive profit of the proposed algorithm is remarkable because it shows that plugging in the EVs does not necessarily cost money when V2G is implemented and it can even bring profit to the EV owners by properly scheduling the control operations.

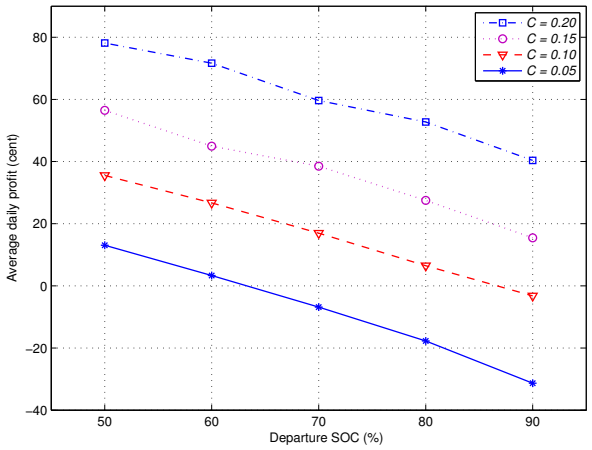


Fig. 4. The average daily profit of the proposed V2G control algorithm using the actual pricing data from PJM under different charging rates C and departure SOCs.

B. Actual Price

We then evaluate our proposed real-time V2G control algorithm using the actual electricity price data from PJM starting from 1:00 January 1st 2010 to 24:00 December 31st 2010 under different charging rates and departure SOCs. The simulation parameters are the same as in the simulated price scenario. For each set of the simulation run, 50 trails of the proposed algorithm are performed.

The simulation results of the proposed algorithm when the departure SOC varies from 50% to 90% and the battery charging rate varies from 0.05 to 0.2 are shown in Fig. 4. As we can see from the figure, the average daily profit decreases linearly as the departure SOC increases while the charging rate stays the same. This is due to the fact that more hours are required for charging in order to meet the increasing departure SOC, which costs more money. If we compare the result of different charging rates at the same departure SOC, it can be found that increasing the charging rate can result in an increasing daily profit. It is not hard to understand this result as the required charging hours diminish when the charging rate is increased. Decreasing required charging hours mean that more hours during the parking time can be utilized to sell power and provide regulation service in order to increase the revenue. Therefore, the EV must either be charged/discharged at the maximum rate, or provide regulation service using full capacity in order to maximize the profit.

VI. Conclusion

In this paper, we investigated the V2G control problem under price uncertainty. The problem is formulated as a discrete-time MDP, where the price uncertainty is modeled via a Markov chain with unknown transition probabilities. We proposed an online learning algorithm to solve the problem adaptively using hourly available pricing information. The proposed algorithm differs from previous algorithms in modeling the price uncertainty brought up by the real-time pricing to make the control decisions. We evaluated the proposed algorithm using both the simulated price and the actual price data from PJM. Simulation results showed that our proposed algorithm is able to increase the profit for the EV owners significantly. More advanced statistical models such as hidden Markov models (HMMs) can also be applied within the same framework in the future.

Acknowledgment

This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Institute for Computing, Information and Cognitive Systems (ICICS) at the University of British Columbia.

References

[1] J. Tomic and W. Kempton, "Using fleets of electric-drive vehicles for grid support," *Journal of Power Sources*, vol. 168, no. 2, pp. 2459–2468, Jun. 2007.

[2] W. Kempton and J. Tomic, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of Power Sources*, vol. 144, no. 1, pp. 280–294, Jun. 2005.

[3] C. Guille and G. Gross, "Design of a conceptual framework for the V2G implementation," in *Proc. of IEEE Energy2030*, Atlanta, GA, Nov. 2008.

[4] W. Kempton and J. Tomic, "Vehicle-to-grid power fundmentals: Calculating capacity and net revenue," *Journal of Power Sources*, vol. 144, no. 1, pp. 268–279, Jun. 2005.

[5] W. Kempton, V. Udo, K. Huber, K. Komara, S. Letendre, S. Baker, D. Brunner, and N. Pearre, "A test of vehicle-to-grid for energy storage and frequecy regulation in the PJM system," University of Delaware and Pepco Holdings, Inc. and PJM Interconnect and Green Moutain College, Tech. Rep., 2008.

[6] S. Han, S. Han, and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. on Smart Grid*, vol. 1, no. 1, pp. 65–72, Jun. 2010.

[7] C. Hutson, G. K. Venayagamoorthy, and K. A. Corzine, "Intelligent scheduling of hybrid and electric vehicle storage capacity in a parking lot for profit maximization in grid power transactions," in *Proc. of IEEE Energy2030*, Atlanta, GA, Nov. 2008.

[8] H. Allcott, *Real Time Pricing and Electricity Markets*, Harvard University, Cambridge, MA, Feb. 2009.

[9] S. Borenstein, "The long-run efficiency of real-time electricity pricing," *Energy Journal*, vol. 26, no. 3, pp. 93–116, 2005.

[10] S. Holland and E. Mansur, "The short-run effects of time-varying prices in competitive electricity markets," *Energy Journal*, vol. 27, no. 4, pp. 127–155, 2006.

[11] A. H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[12] A. Conejo, J. Morales, and L. Baringo, "Real-time demand response model," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 236–242, Dec. 2010.

[13] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *Proc. of IEEE SmartGridComm*, Gaithersburg, MD, Oct. 2010.

[14] A. Gonzalez, A. San Roque, and J. Garcia-Gonzalez, "Modeling and forecasting electricity prices with input/output hidden Markov models," *IEEE Trans. on Power Systems*, vol. 20, no. 1, pp. 13–24, Feb. 2005.

[15] A. Brooks, "Vehicle-to-grid demonstration project: Grid regulation ancillary service with a battery electric vehicle," AC Propulsion, Inc., Tech. Rep., 2002.

[16] T. Gage, "Final report: Development and evaluation of a plug-in HEV with vehicle-to-grid power flow," AC Propulsion, Inc., Tech. Rep., 2003.

[17] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, 1998.

[18] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, 2007, vol. 2.

[19] V. Borkar and S. Meyn, "The ODE method for convergence of stochastic approximation and reinforcement learning," *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 447–469, Jan. 2000.

[20] PJM Interconnection, http://www.pjm.com/.