A Leader-Follower Computational Learning Approach to the Study of Restructured Electricity Markets: Investigating Price Caps

Kurian Tharakunnel and Siddhartha Bhattacharyya
Department of Information and Decision Sciences
University of Illinois at Chicago
kthara1@uic.edu, sidb@uic.edu

Abstract

This paper discusses the use of a computational learning approach based on a leader-follower multiagent framework in the study of regulation of restructured electricity markets. In a leader-follower multiagent system (LFMAS), a leader (regulator) determines an appropriate incentive, which motivates a set of self-interested followers (the generators, in this case) to act such that some measure of overall performance is maximized. In the computational learning approach presented, models of followers as well as the leader incorporate reinforcement learning, allowing the exploration of outcomes with different incentives, and also the learning of 'optimal' incentive given some measure of desired overall performance. The approach is demonstrated in studying the effect of price caps on the outcome of electricity auctions (uniform and discriminatory) in oligopoly settings for which analytical treatments do not exist.

1. Introduction

Search for effective market and non-market mechanisms that address market weaknesses such as market power associated with the operation of restructured electricity markets has been on ever since the restructuring efforts were undertaken by the various governments. This has motivated several studies that examined the various facets of this issue employing a variety of methods including analytical, empirical, and simulation.

Several features unique to the electricity markets such as inelasticity of demand and inability to store electricity make their study a complex task. Thus, most studies on electricity markets rely on stylized models that abstract away several of these complexities and focus on one or more of the issues. A promising recent approach that allows for incorporating better realism in electricity market models is the agent-based computational learning approach. In an agent-based model, the electricity market is modeled as a multiagent system consisting of autonomous agents with learning capabilities. Several studies in the past have employed this approach to the study of electricity markets [3], [4]. However, most agent-based computational studies of electricity markets focus on the market design aspects and do not include non-market mechanisms like price caps. Studies of restructured electricity markets in general have ignored the non-market mechanisms or have failed to consider them in a comprehensive fashion [10].

In this work we present a multiagent model for the study of non-market mechanisms (also known as regulatory mechanisms) in restructured electricity markets. From a multiagent perspective, there is a fundamental difference between the structure of a market design problem and that of a regulation problem. In a market design problem, the strategic interaction is amongst a set of generators that have similar roles and objectives. On the other hand, in a regulation problem, a regulator interacts with a set of competing generators. While the individual generators have profit maximization as the objective, the regulator's objective is often maximization of some measure of overall performance such as social welfare. Regulation problems thus have a hierarchical structure.

We present a multiagent model that is especially suited for the study of regulation problems. In this model, the regulated market is

1530-1605/08 $25.00 © 2008 IEEE

modeled as a leader-follower multiagent system (LFMAS). This model has a leader-agent representing the regulator that designs the regulatory input, and follower-agents representing generators who take this input into consideration while determining their own actions. Followers are self-interested, and the leader's regulatory action provides an incentive that affects the followers' payoffs from different actions, such that their combined actions lead to the maximization of some measure of overall performance (e.g., social welfare). Leader-follower problems [1] are models of hierarchical decision problems with applications in regulation and control. For example, Keyhani [11] proposed a leader-follower framework for the control of electricity markets.

In the computational learning approach that this paper presents, both the leader and the followers' models incorporate reinforcement learning. The followers jointly learn to act such that their own self-interested goals are maximized, while the leader's learning seeks the optimal regulatory control with respect to the leader's goal. We show how this approach can be useful by examining some recent studies pertaining to price-cap regulation in electricity markets.

Most restructured electricity markets in operation use price caps as part of their regulatory mechanism. For example, Pennsylvania-New Jersey-Maryland (PJM) market has elaborate rules for setting price caps in its operation. While price caps are pervasive in electricity markets, there are differing views about their effectiveness as a regulatory tool. Price caps have been employed as a tool for preventing excessive bidding by generating companies. Theoretically, an appropriately set price cap is shown to lower prices and increase aggregate output from the firms in the market [2]. On the other hand, it has been shown that use of price caps, in the long run, may result in disincentive for investment and thus may lead to shortage of capacity [10]. We emphasize here that in this paper our focus will be in the short run effects of price caps in controlling market power.

We consider price caps in the context of markets operating on uniform or discriminatory (pay-as-bid) auction formats. Specifically, we examine the effect of price caps when the generators face an uncertain demand. A random demand can occur due to uncertain events like plant or transmission line outages, extreme weather, etc. and is especially relevant in day-ahead markets where the suppliers submit bids that remain valid for 24 hours [7]. A recent study [6] showed that under certain assumptions, the results about price caps in the certainty case do not hold in general when the demand is uncertain. Specifically, using a Cournot model of the market this study showed that the average prices might not be non-decreasing in price caps under the presence of demand uncertainty. We examine this phenomenon in the context of uniform and discriminatory auctions in an oligopolistic setting. The results from our experiments did not indicate this behavior in the auction settings we considered. However, our results show that with uncertain demand, the strategic behavior of generators in a oligopolistic setting are quite different from that in a duopoly setting in both auction formats.

The rest of the paper is organized as follows. The leader-follower multiagent model is introduced in Section 2 and the computational learning approach based on this model is described in Section 3. Section 4 describes electricity auctions. Section 5 presents the results from experiments using the proposed computational learning approach. Discussion of contributions, limitations of this study, and future work appear in Section 6.

2. Leader-Follower Multiagent Model

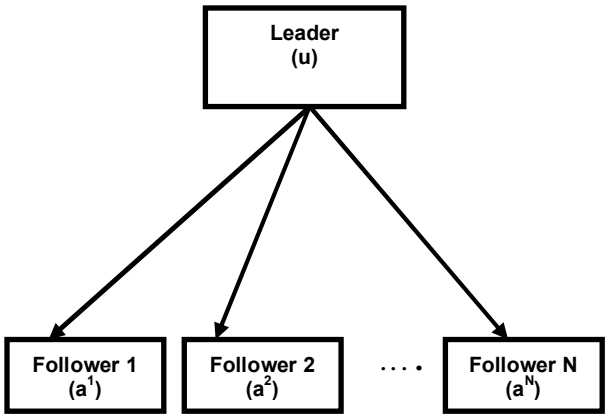

Figure 1. A Leader-Follower Multiagent System (LFMAS)

A Leader-Follower Multiagent System (LFMAS) (Figure 1) consists of a single leader agent and N, $(N>1)$ follower agents. Let $u \in U \subset \Re$ and $a^n \in A^n \subset \Re, n = 1, 2, \cdots, N$ are the actions available to the leader and the followers respectively. The leader and the

followers make their decisions sequentially, with the leader making the decision first and announcing it to the followers. The followers, after knowing leader's decision, make their individual decisions concurrently and competitively. In other words, the followers play a Nash game after knowing the leader's decision. The payoffs of the leader and the followers are interrelated in the sense that the followers' payoffs are contingent on the leader's decision while the leader's payoff is a function of the followers' actions. In game theory, the above-described strategic interaction between the leader and the followers is known as a Stackelberg game and the associated equilibrium solution is known as a Stackelberg equilibrium [1].

Let the payoff functions of the leader and the followers be $V^l(u,a^1,a^2,\cdots,a^N)$, and $V^{f1}(u,a^1,a^2,\cdots,a^N)$, $V^{f2}(u,a^1,a^2,\cdots,a^N)$ $\cdots$, $V^{fN}(u,a^1,a^2,\cdots,a^N)$ respectively. For a given incentive u announced by the leader, let the N-tuple $(a_u^{1*},a_u^{2*},\cdots,a_u^{N*})$ be the unique Nash equilibrium for the followers' subgame such that

$$V^{f1}(u,a_u^{1*},a_u^{2*},\cdots,a_u^{N*}) \geq V^{f1}(u,a^1,a_u^{2*},\cdots,a_u^{N*}), \forall a^1 \in A^1$$

$$V^{f2}(u,a_u^{1*},a_u^{2*},\cdots,a_u^{N*}) \geq V^{f1}(u,a_u^{1*},a^2,\cdots,a_u^{N*}), \forall a^2 \in A^2$$

$$\vdots$$

$$V^{fN}(u,a_u^{1*},a_u^{2*},\cdots,a_u^{N*}) \geq V^{f1}(u,a_u^{1*},a_u^{2*},\cdots,a^N), \forall a^N \in A^N$$

The leader's problem is to determine the optimal incentive u^* such that

$$u^* = \arg\max_u V^l(u,a_u^{1*},a_u^{2*},\cdots,a_u^{N*}) \qquad (1)$$

The $(N+1)$ tuple $(u^*,a_{u^*}^{1\,*},a_{u^*}^{2\,*},\cdots,a_{u^*}^{N*})$ is then called a Stackelberg equilibrium. It is assumed that the followers' subgame has a unique Nash equilibrium for every incentive decision announced by the leader.

The above approach to the solution of Stackelberg games is analytically intractable when there are more than two followers. Further, the above solution approach assumes that the leader knows the payoff functions of every follower and can compute the corresponding equilibrium of the followers' subgame for each of it's actions and thereby arrive at its optimal decision. Also, the followers are assumed to have common knowledge of their payoff functions to play the corresponding equilibrium actions. This approach thus puts strong assumptions about the informational and computational capabilities of the leader and the followers. In contrast, the computational learning approach we propose in this work assumes limited information requirements for the players. In particular, we assume that the leader and the followers observe only the rewards they receive for their actions. The next section presents this approach.

3. Computational Learning Approach

In the computational learning approach we use, the leader and the followers are adaptive learners that use simple reinforcement learning (RL) schemes to learn their respective optimal strategies- leader learning the optimal incentive, and the followers learning their corresponding equilibrium responses.

3.1. Reinforcement learning preliminaries

Reinforcement learning [15] is a popular model of learning frequently employed in agent-based models to represent agent behavior. In this model of learning, agents use rewards received from past actions to learn optimal actions. One of the most successful and popular RL schemes is the Q-learning proposed by Watkins [16]. In what follows we use a very simple form of Q-learning to describe the basic approach of RL.

Q-learning uses a set of quantities (one for each admissible action of the agent) called Q values which are basically estimates of the expected rewards for different actions. The algorithm starts with some initial estimates for Q values. Subsequently, at every time step t, the Q value corresponding to the current action a is updated as follows:

$$Q_{t+1}(a) = Q_t(a) + \lambda_t\left(R(a)_t - Q_t(a)\right)$$

where $R(a)_t$ is the reward received for taking action a at time t, and $\lambda_t (0 < \lambda_t < 1)$ is the learning rate that controls the magnitude by which the Q-values are modified at each updating step. The learning rate is gradually decayed so that $\sum_t \lambda_t = \infty$ and $\sum_t \lambda_t^2 = 0$.

It is proved that (when Q-values are stored in a tabular format) Q-learning converges to the optimal Q-values under the condition that each admissible action is performed in each state infinitely often in infinite number of decision epochs. In practice, this condition is met by implementing an *explore/exploit* action selection scheme for the agent. Under this scheme, at every time step, the agent selects a random action with some small probability. One widely used technique for this purpose is the Boltzmann action selection scheme [15]. Under this scheme, at any time step, the probability of selecting action a when the state is s is $\frac{e^{Q(s,a)/T}}{\sum_a e^{Q(s,a)/T}}$ where $Q(s,a)$ is the current estimate of Q-value for state-action pair (s,a) and T a "temperature" parameter that controls the degree of randomness in action selection. The temperature T is gradually reduced from a predetermined maximum to a predetermined minimum by an appropriate decaying scheme.

There have been several attempts to extend the Q-learning approach to multiagent systems [4],[9], and [13]. These efforts focus on RL schemes for multiagent systems where the agents have symmetrical roles and the associated game theoretic solution concept is a Nash equilibrium. In contrast, the RL approach we present in this work is for leader-follower multiagent systems where the agents have asymmetric roles and as shown in the previous section, the associated game theoretic solution concept is a Stackelberg equilibrium.

3.2. An RL approach for LFMAS

The proposed RL approach consists of leader's learning scheme modeled as a single agent RL and followers' learning scheme modeled as a multiagent RL. A crucial point though is, the coupling of these two learning processes so that condition (1) is achieved. This coupling is obtained by making the leader's learning process learn at a slower rate than the followers' learning process. We assume that the leader makes a new decision every m period while the followers repeatedly play the game for m periods. The leader's learning scheme then resembles the single agent Q-learning except for the fact that the immediate reward for the leader is the reward accrued over m periods.

The followers' learning scheme is an adaptation of a Q-learning algorithm for repeated games proposed recently by Leslie and Collins [12]. There are two important features that make this algorithm especially attractive for our purposes. First, this algorithm uses player dependent learning rates for the update of Q-values of individual agents, which under certain assumptions have good convergence properties. Secondly, the algorithm uses a smooth best response (SBR) [8] scheme for action selection that enables the agents to learn mixed strategies. This is important in many applications especially the regulation problem we address in this work where the generators have a mixed strategy equilibrium.

The original Q-learning algorithm, being an optimum-seeking algorithm, can learn only pure strategies. The use of SBR action selection addresses this issue. A SBR scheme maintains a positive probability for every action in an agent's action set to get selected. One way to implement an SBR action selection scheme is to use a Boltzmann action selection scheme described earlier with the temperature parameter held constant at a very small value. When T is held constant, every action in the action set will always have a positive probability (though small) of getting selected. [12] show that, with smooth best responses, an agent's strategy converges towards a Nash distribution which is an approximation of the mixed strategy Nash equilibrium of the game.

The proposed RL algorithm for LFMAS is as follows.

1. Leader starts with an incentive u.
2. Followers play a game

each follower n selects an action a^n according to SBR scheme

each follower n receives a reward r^n and updates its Q value Q^n using the following update scheme

$$Q^n(u,a^n) \leftarrow Q^n(u,a^n) + \lambda^n \left[r^n - Q^n(u,a^n) \right]$$

3. Step 2 is repeated m times
4. Leader receives the aggregate reward r^l since the last incentive decision and updates its Q value Q^l using the following update scheme

$$Q^l(u) \leftarrow Q^l(u) + \lambda^l \left[\frac{r^l}{m} - Q^l(u) \right]$$

5. If the termination criterion is not met the leader selects an incentive u using explore/exploit action selection scheme

Return to Step 2.

The parameter m decides the number of times the followers play the game before the leader updates its Q value. Note that, the update of the leader and the followers happen at different time scales. Follower Q-values are updated in every time period whereas for the leader, the update happens only once in m periods. The followers' maintain separate set of Q-values for each incentive decision of the leader.

The player dependent learning rates for the followers' learning is implemented as follows. The learning rate for follower n at time step t is set as $\lambda_t^n = (t + C)^{-\theta^n}$ where $\theta^n \in (0.5, 1]$ and C a constant. By selecting θ^n differently for each follower, the required sequence of learning rates is obtained.

Each follower agent employs a SBR action selection scheme that uses the Boltzmann function with the "temperature" parameter T set at a very small value. As discussed earlier, this enables the followers to learn mixed strategies. We also use the Boltzmann function for explore/exploit action selection of the leader, but with a decaying T as in single-agent Q-learning.

4. Electricity auctions

Most restructured wholesale electricity markets use auctions as the primary market mechanism for electricity trading. While there are several forms of auctions in existence, the most widely used format in electricity trading is the *uniform auction*, also known as first-price auction. However, some markets, like for example, England and Wales, have recently switched to a *discriminatory auction* (also known as pay-as-bid) format and several others are considering this change. A recent debate in the study of restructured electricity markets is the relative merits and drawbacks of these two auction formats in electricity trading. While our interest in this paper is not to compare these two auction formats, because of the importance of these two forms of auctions in electricity trading, we undertake our study using these two auction formats.

In both auction formats, the suppliers submit their bids for a period by specifying the minimum offer prices and available capacities at those prices. The auctioneer considers these bids and the forecasted demand for the period and then decides dispatch quantities for the period by allocating the least cost supplier first until all demand is met. The main difference between the two formats is in the payments to the suppliers. In the uniform auction, all allocated units are paid at the price of the marginal accepted unit. In discriminatory auctions, the suppliers are paid at their offer prices for the quantities allocated.

Several studies have analyzed uniform and discriminatory auctions in the context of restructured electricity markets. Son et al. [14] compares the equilibrium characteristics of uniform and discriminatory auctions in a duopoly setting. A more recent study by Fabra et al. [7] examines bidding behavior and market outcomes in uniform and discriminatory auctions under a variety of market situations. However, most of these studies are limited to duopoly situations as oligopoly settings are analytically hard to solve.

Our focus in this study is on the effect of price caps on the market outcome when the auction format used is uniform or discriminatory. The next section discusses the results from experiments using the computational learning approach discussed in Section 3.

5. Experiments and results

In these experiments we consider a market with multiple number of generators operating on either uniform or discriminatory auction formats. The demand faced by the market is uniformly distributed. As in [7] we assume that generators offer their entire capacity for bid. While this is a simplification, it turns out that this may be a reasonable assumption in many situations. For example, a recent study of bidder behavior at New York ISO [17] shows that generators submit most of their installed capacity once they choose to bid into market. We implemented the agent model using Swarm 2.2 simulation kit.

5.1. Strategic behavior under demand uncertainty

We begin by studying the equilibrium-bidding behavior of the generators under duopoly and oligopoly settings with a fixed price cap (the leader's learning turned off). In the duopoly case we consider one large capacity generator with capacity 15 and a fixed marginal cost of 3, and a small capacity generator with capacity 5 and marginal cost 4. The demand in this case is uniformly distributed in the range

$[6,15]$. In the oligopoly setting there are five generators with one large capacity generator, three small capacity generators, and one medium capacity generator. The capacities and costs of large and small capacity generators are as in the duopoly case whereas the medium capacity generator has a capacity of 10 and marginal cost 3.5. The demand in this case varies in the range $[15,30]$. The price cap is set at 12 in all cases.

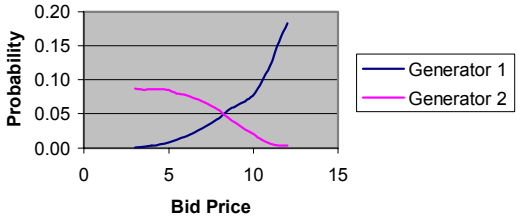

Figure 2. Bidding strategies in uniform auction (duopoly)

Figure 2 shows the bidding strategies of the generators in duopoly case under uniform auction. As expected, both generators employ mixed strategies with the large capacity generator more often bidding closer to the price cap and the smaller capacity generator bidding closer to its marginal cost. Figure 3 shows the bidding strategies of the generators in the oligopoly case under uniform auction. In this case all generators have mixed strategies that specify bidding closer to their respective marginal costs. This is because, as the number of generators increase, the competition forces the generators to bid closer to their marginal costs. Figure 4 and 5 show the bidding strategies of generators in the duopoly and oligopoly settings respectively under discriminatory auction format.

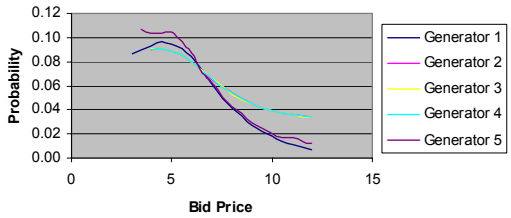

Figure 3. Bidding strategies in uniform auction (oligopoly)

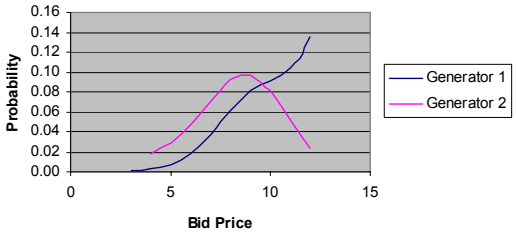

Figure 4. Bidding strategies in discriminatory auction (duopoly)

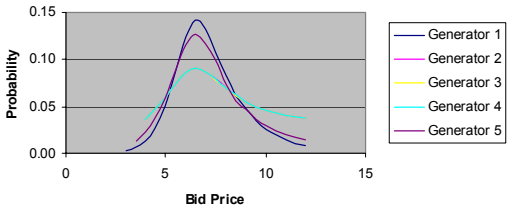

Figure 5. Bidding strategies in discriminatory auction (oligopoly)

5.2. Effect of price cap

In these experiments we let the leader to vary the price cap setting. Figure 6 shows the plot of cost of supply for different price caps under unform and discriminatory auction settings in duopoly while Figure 7 shows the same in oligopoly. It can be observed that the cost of supply is non-decreasing in the price cap in all cases here. This is in contrast to the recent findings that average price may not be non-decreasing in the price cap under demand uncertainty. Another interesting observation is that the cost of supply under discriminatory auction is lower than the cost of supply in uniform auction under different price cap settings both in duopoly and oligopoly. This is consistent with the analytical results of [7] and [14].

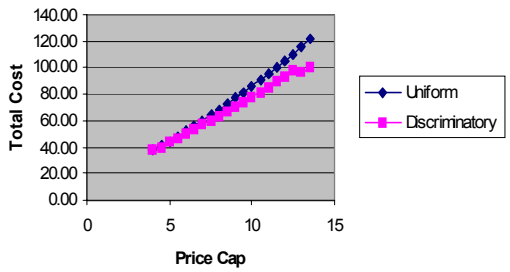

Figure 6. Cost of supply under different price caps (duopoly)

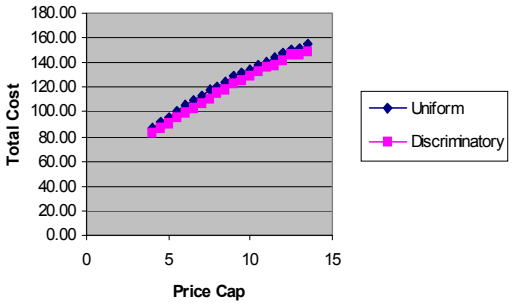

Figure 7. Cost of supply under different price caps (oligopoly)

6. Discussion

The main contribution of this paper is a new multiagent model and the associated computational learning approach for the study of regulation problems. The computational approach is very general and can be employed to study a variety of regulation problems not only in electricity markets but also in many other regulation situations such as emission control, pricing communication networks etc.

The results in this study provide interesting observations about bidder strategies and price cap effectiveness in oligopolistic settings of electricity auctions, which are not easily amenable to theoretical analysis. In future work we would like to extend this study to cases that include elastic demand and multiple bids. Another interesting avenue for further work would be to employ this approach to model an existing market using actual market parameters.

One weakness of the present study is the unknown convergence properties of the computational learning approach. Establishing convergence conditions of the learning scheme presented is another important future work to be undertaken.

7. Acknowledgment

This work was supported by National Science Foundation grant ECS-0601590

8. References

[1] Basar, T., & Olsder, G. J. (1995). Dynamic noncooperative game theory. London: Academic Press.

[2] Borenstein, S. (2002). The trouble with electricity markets: Understanding california's restructuring disaster. Journal of Economic Perspective, 16 (1), 191 211.

[3] Bower, J., & Bunn, D. (2001). Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the england and wales electricity market. Journal of Economic Dynamic Control , 25 , 561-592.

[4] Bowling, M., & Veloso, M. (2002). Multiagent learning using a variable learning rate. Artifcial Intelligence, 136 , 215-250.

[5] Bunn, D. W., & Oliveira, F. S. (2003). Evaluating individual market power in electricity markets vis agent-based simulation. Annals of Operations Research, 121 , 57-77.

[6] Earle, R., Schmedders, K., & Tatur, T. (2007). On price caps under uncertainty. Review of Economic Studies, 74 , 93-111.

[7] Fabra, N., von der Fehr, N. H., & Harbord, D. (2006). Designing electricity auctions. The Rand Journal of Economics, 37 (1), 23-46.

[8] Fudenberg, D., & Levin, D. K. (1998). The theory of learning in games. Cambridge, MA: MIT Press.

[9] Hu, J., & Wellman, M. P. (2003). Nash q-learning for general-sum stochastic games. Journal of Machine Learning Research, 4 , 1039-1069.

[10] Joskow, P., & Tirole, J. (2006). Reliability and competitive electricity markets. The Rand Journal of Economics (to appear)

[11] Keyhani, A. (2003). Leader-follower framework for control of energy services. IEEE Transactions on Power Systems, 18 (2), 837-841.

[12] Leslie, D. S., & Collins, E. J. (2004). Individual q-learning in normal form games. Submitted to SIAM J. of Control and Optimization.

[13] Littman, M. (1994). Markov games as a framework for multi-agent reinforcement learning. In Proceedings of the Eleventh International Conference on Machine Learning pp. 157-163

[14] Son, Y. S., Baldick, R., Lee, K.-H., & Siddiqi, S. (2004). Short-term electricity market auction game

analysis: Uniform and pay-as-bid pricing. IEEE Transactions on Power Systems, 19 (4), 1990-1998.

[15] Sutton, R. S., & Barto, A. (1998). Reinforcement learning: An introduction. Cambridge, MA: MIT Press.

[16] Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. Machine Learning, 8 (3/4), 279-292.

[17] Zhang, N., Mount, T., & Boisvert, R. (2007). Generators' bidding behavior in the NYISO day-ahead wholesale electricity market. In 40th Hawaii International Conference on System Sciences (HICSS-2007) (pp. 11-17). Hawaii.