

Modeling of Suppliers' Learning Behaviors in a Market Environment

Nanpeng Yu*

Chen-Ching Liu*, *Fellow, IEEE*

Leigh Tesfatsion**, *Member, IEEE*

*Department of Electrical and Computer Engineering

**Department of Economics
Iowa State University

Abstract: An important objective of electricity suppliers is to maximize their profits over a planning horizon and comply with the market rules. This objective requires suppliers to learn from their bidding experience and behave in an anticipatory way. With volatile Locational Marginal Prices (LMPs), ever-changing transmission grid conditions, and incomplete information about other market participants, decision making for a supplier is a complex task. A learning algorithm that does not require an analytical model of the complicated market but allows suppliers to learn from experience and act in an anticipatory way is a suitable approach to this problem. Q-Learning, an anticipatory reinforcement learning technique, has all these desired properties. Therefore, it is used in this research to model the learning behaviors of electricity suppliers in a Day-Ahead electricity market. The Day-Ahead electricity market is modeled as a multi-agent system with interacting agents including supplier agents, Load Serving Entities and a Market Operator. Simulation of the market clearing results under the scenario in which agents have learning capabilities is compared with the scenario where agents report true marginal costs. It is shown that, with Q-Learning, electricity suppliers are making more profits compared to the scenario without learning due to strategic gaming. As a result, the LMP at each bus is substantially higher.

Keywords: Electricity Market, Supplier Modeling, Competitive Markov Decision Process, Q-Learning.

1. Introduction

Strategic bidding is an important issue in the wholesale electricity market. Electricity prices change as a result of transmission network congestion, which may be caused by strategic bidding or heavy load. For PJM, the total congestion costs were \$750 million in 2004 and \$2.09 billion in 2005. Learning may also allow larger electricity suppliers to use their market power and bid strategically. In California [1], electricity expenditure in the wholesale market increased from \$2.04 billion in the summer of 1999 to \$8.98 billion in the summer 2000. It is estimated that 59% of this increase was due to increased market power. Learning to bid in the wholesale market is also crucial for smaller electricity suppliers who have a desire to recover the cost of their investment in generation by avoiding over or under-bidding. Research on the learning behavior of

electricity suppliers will provide insights into gaming on the market and the power grid. This may allow market designers to develop appropriate market rules to discourage strategic bidding and enhance market efficiency.

Researchers have used various learning methods to model electricity suppliers' behavior. The learning configuration for suppliers in [2] is a version of a stochastic reactive reinforcement learning developed by Alvin Roth and Ido Erev. In this configuration, agents have finite fixed action domains, are backward looking, and rely entirely on response learning. Average reward γ -greedy reinforcement learning was used in [3] to model the learning and bidding processes of suppliers. With this scheme, each supplier uses greedy selection as its action choice rule with probability $(1 - \gamma)$, and random action selection with probability γ . Thus, γ determines the trade-off between exploitation of available information and exploration of untested actions. The trading agents modeled in [4] use GP-Automata to compute their bidding strategies for the current market conditions. Finally, in the area of multi-agent reinforcement learning, Nash Q-Learning [5] was designed specifically as a potential technique to represent agents' learning behavior in a multi-agent context.

This paper is focused on how to model electricity suppliers' learning behavior by Q-Learning. In addition, load serving entities that have demand-side response are considered in this multi-agent market environment.

The remainder of this paper is organized as follows. Section 2 describes the configuration of the Day-Ahead Market. Section 3 briefly discusses the problem faced by supplier agents and presents a novel Q-Learning model for their learning behavior. Section 4 presents numerical results for a 5-node sample system. Finally, conclusions and suggestions for the future work are given in section 5.

2. Day-Ahead Market Model

The Day-Ahead electricity market is modeled as a multi-agent system with three types of agents interacting with one another. These agents are supplier agents, Load Serving Entities (LSEs), and a market operator (MO). On the morning of day D supplier agents submit supply offers and LSEs submit demand bids for the Day-Ahead Market to the MO. During the afternoon, the MO runs a market-clearing

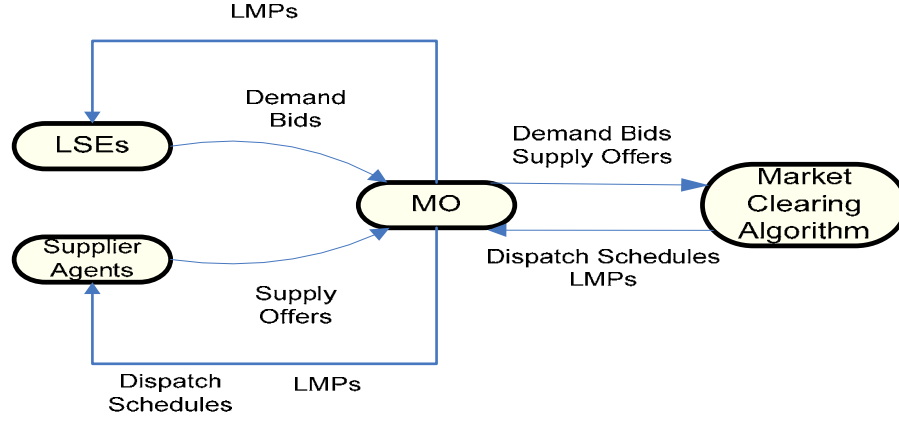


Fig. 1: Multi-Agent Day-Ahead Market Environment

algorithm (similar to an optimal power flow) to match supply to demand and determine dispatch schedules and LMPs. At the end of the process, the MO sends the dispatch schedules and LMPs to the supplier agents and LSEs for day D+1. The interaction among the MO, LSEs and supplier agents is shown in Fig. 1.

2.1 Load Serving Entity Model

LSEs purchase bulk power from the Day-Ahead market to serve load. Without loss of generality, it is assumed that LSEs do not have generation units and one LSE only serves load at one location in the power system. Suppose that the number of LSEs in the Day-Ahead market is J . On day D, LSE j submits a load profile for day D+1. This load profile specifies 24 hours of MW power demand $P_{L_j}(H)$, $H=0, 1, \dots, 23$.

It is assumed that demand-side response is available to LSEs. The demand-side response works as follows. If the day D peak hour LMP, $LMP_{L_j}(H_{peak})$, at the bus where LSE j is serving load is higher than a critical value, then it reduces its peak hour demand for day D+1 by 2%. If this LMP does not exceed the critical value, it will not curtail its peak hour demand. Therefore, each LSE has two states. If the LMP at its node is below the critical value, it is in state 0, i.e., $S_{L_j} = 0$, and it will submit a normal load profile for day D+1. If the LMP at its node is above the critical value, it is in state 1, i.e., $S_{L_j} = 1$, and it will submit a curtailed load profile for day D+1.

2.2 Supplier Agent Model

Supplier agents sell bulk power to the Day-Ahead market. For simplicity, it is assumed that one supplier agent only has one generation unit. However, this model can be extended to one supplier with multiple generation units. Suppose the number of supplier agents in the Day-Ahead market is I , and the MW power output of generator i in

same hour H is p_{Gi} . Generator i has lower and upper limits denoted by p_{\min_i} and p_{\max_i} for its hourly MW power output. For generator i , the hourly total production cost $C_i(p_{Gi})$ for production level p_{Gi} is represented in a quadratic form:

$$C_i(p_{Gi}) = a_i \cdot p_{Gi} + b_i \cdot p_{Gi}^2 + F_i \quad (1)$$

where a_i, b_i and F_i (pro-rated fixed cost) are given constants. By taking derivatives on both sides of (1), the marginal cost function for Generator i is obtained, i.e.,

$$MC_i(p_{Gi}) = a_i + 2 \cdot b_i \cdot p_{Gi} \quad (2)$$

On each day D, the supplier agents submit to the Day-Ahead market a supply offer for day D+1 that includes two components. The first component is its reported marginal cost function given by:

$$MC_i^B(p_{Gi}) = a_i^B + 2 \cdot b_i^B \cdot p_{Gi} \quad (3)$$

The second component is the hourly MW power output upper limits, denoted by $p_{\max_i^B}$. Suppose, on day D, supplier agents submit their supply offers for day D+1 to the MO, and the market clearing program calculates LMPs and dispatch schedules. Let $LMP_{G_i}(H)$ denote the LMP of hour H at the bus where supplier i 's generation unit is located, and $p_{Gi}^*(H)$ denote the MW power output in the dispatch schedule posted by the MO. Supplier agent i 's profit on day D is obtained by summing 24 hours' of profits on that day:

$$\pi_{iD} = \sum_{H=0}^{24} [p_{Gi}^*(H) \cdot LMP_{G_i}(H) - C_i(p_{Gi}^*(H))] \quad (4)$$

The accumulated profit of generator i on day D is given by:

$$AP_i(D) = AP_i(D-1) + \pi_{iD} \quad (5)$$

2.3 Market Operator Model

The MO for this Day-Ahead market is responsible for clearing the market based on the information submitted by

LSEs and supplier agents. The MO uses a market clearing algorithm to determine the LMP at each bus and MW power output for each generation unit at each hour. Since only MW power is considered in this model, a DCOFP problem can be formulated as follows:

$$\min \sum_{i=1}^I (a_i^B \cdot p_{Gi} + b_i^B \cdot p_{Gi}^2) \quad (6)$$

Subject to

$$P_k - P_{gk} + P_{dk} = 0, \quad k = 1, \dots, N_b \quad (7)$$

$$|H\delta| \leq F_{\max} \quad (8)$$

$$p \min_i^B \leq p_{Gi} \leq p \max_i^B, \quad i = 1, \dots, I \quad (9)$$

where N_b denotes the total number of buses in the system,

P_k represents the net power injection at bus k , P_{gk} denotes the total MW power generation at bus k , P_{dk} is the total MW demand at bus k , H denotes the line flow matrix, δ denotes the voltage angle difference vector, and F_{\max} is the vector of maximum line flows.

The objective of the DCOFP is to minimize the total variable generation cost based on supplier offers and LSE bids. The constraints are MW power balance constraints for each bus $k = 1, \dots, N_b$, MW thermal limit constraints for each line, and MW production limits for each generator $i = 1, \dots, I$. The DCOFP program of MATPOWER [6] is used in this research.

3. Model for Suppliers' Learning Behavior

Many learning methods could be used to model a supplier's behavior, including Reactive Reinforcement Learning, Belief-Based Learning, Q-Learning, Genetic Algorithms, and Artificial Neural Networks. Q-Learning is judged to be a most promising technique for modeling of the electricity suppliers' learning behavior. Reactive Reinforcement Learning relies entirely on backward-looking stimulus and response learning, which prevents agents from perceiving any way to sacrifice immediate reward for long-term benefits. Belief-Based Learning assumes that agents use estimates of opportunity costs to select the best action. However, the opportunity cost of the actions chosen is hard to measure, unless one has complete information about what actions other agents have chosen, which is not true in the electricity market. The result of a Genetic Algorithm depends heavily on the fitness function which is difficult to specify in the context of electricity markets where the total discounted reward is to be maximized. To ensure successful use of Artificial Neural Networks, a large amount of training data is needed which may not be available. Due to the model-free and anticipatory properties that Q-Learning possesses, it appears to be an attractive learning method for the proposed application.

Q-Learning, developed by Watkins [7], is a form of anticipatory reinforcement learning that allows agents to learn how to act in a controlled Markovian domain. A controlled Markovian domain implies that the environment is Markovian in the sense that state transition probabilities from state x to state y only depends on x , y and the action a taken by the agent, and not on other historical information. It works by successively updating estimates for the Q-values of state-action pairs. The Q-value $Q(x, a)$ is the expected discounted reward for taking action a at state x and following an optimal decision rule thereafter. Once these estimates have converged to the correct Q-values, the optimal action in any state is the one with the highest Q-value.

By the procedure of Q-Learning, in the n^{th} step the agent observes the current system state x_n , selects an action a_n , receives an immediate payoff r_n , and observes the next system state y_n . The agent then updates its Q-values estimates using a learning parameter α_n and a discount factor γ [7]:

If $x = x_n$ and $a = a_n$,

$$Q_n(x, a) = (1 - \alpha_n)Q_{n-1}(x, a) + \alpha_n[r_n + \gamma V_{n-1}(y_n)] \quad (10)$$

Otherwise

$$Q_n(x, a) = Q_{n-1}(x, a) \quad (11)$$

$$\text{where } V_{n-1}(y) \equiv \max_b \{Q_{n-1}(y, b)\} \quad (12)$$

It is proven by Watkins in [8] that if (1) the state and action-values are discrete, (2) all actions are sampled repeatedly in all states, (3) the reward is bounded, (4) the environment is Markovian and (5) the learning rate decays appropriately, then the Q-value estimates converge to the correct Q-values with probability 1.

In a multi-agent context such as the Day-Ahead market model presented in this paper, the system may not be Markovian because state transition probabilities may depend on actions taken by other agents. Therefore, there is no guarantee that Q-Learning will converge to the correct Q-values.

A Generation Company (GENCO) usually has several generation plants located at different buses of the system. For simplicity, Q-Learning is used to model electricity suppliers that are assumed to have only one generation unit. Nevertheless, by a similar approach, Q-Learning could be implemented for supplier agents with multiple generation units at different locations.

A novel approach to the implementation of Q-Learning for a supplier agent is presented here. The supplier agent views the Day-Ahead market as a complex system with different

states. The system state on day D , X^D , is defined as a vector for the states of all LSEs. Hence the state vector on day D can be expressed as $X^D = \{S_{L1}, S_{L2}, \dots, S_{LJ}\}$, where J is the number of LSEs. The cardinality of the state space is 2^J since each LSE has two states, i.e., reduced peak load or not based on demand-side response. Electricity suppliers might have market power. Thus, it is assumed that supplier agents are capable of forecasting the LSEs' states. In other words, the state vector is predictable by the supplier agents.

The action domain of supplier agent i , AD_i , is defined as a vector of bidding information. This vector consists of the marginal cost function parameters a_i^B and $2 \cdot b_i^B$, and the hourly MW output upper limit $p \max_i^B$. The cardinality of the action domain, $M^a \times M^b \times M^{\max}$, is given by the product of the number of possible a_i^B , $2 \cdot b_i^B$ and $p \max_i^B$ values.

Consider the beginning of each day D . A supplier agent first makes a prediction of the system state, which is represented by x . It next chooses an action according to a Gibbs/Boltzmann probability distribution, i.e.,

$$p_D(a) = \frac{e^{Q(x,a)/T_D}}{\sum_{b \in AD_i} e^{Q(x,b)/T_D}} \quad (13)$$

where T_D , which depends on D , is a temperature parameter that models a decay over time.

Having chosen an action a , the supplier agent will submit its supply offer to the MO. Once the market is cleared, the supplier agent will receive its reward, which is the profit for day $D+1$. Then the agent will use this reward to update its Q-value estimates according to equations (10) to (12). The Q-value estimates of an agent are said to have converged if under all states x the agent chooses some action with a probability exceeding 0.99. If the Q-value estimates of all the agents have converged, the simulation terminates.

The parameters that are used to implement the Q-Learning algorithm are set in the following way:

Discount factor $\gamma = 0.7$

Learning parameter α for a state-action pair (x, a) is set to

be $\alpha = \frac{1}{T_{(x,a)}^\omega}$, where $T_{(x,a)}$ is the number of time that

action a has been taken in state x .

$\omega = 0.77$

The temperature parameter T_D is given by: $\frac{1}{T_D} = 1.7 \times 10^{-9} \times (D)^6$, where D is the number of days

that have currently been simulated.

The cardinality of the action domain is $M^a \times M^b \times M^{\max} = 4 \times 4 \times 4$, in which a_i^B and b_i^B range from 1 to 3 times of their true values, $p \max_i^B$ ranges from 97% to 100% of the true upper limit.

4. Numerical Study

4.1 Test Case

The 5-bus transmission grid used here for simulation is taken from ISO-NE/PJM training manuals, where it is used to illustrate the determination of Day-Ahead market LMP solutions. A one-line diagram of the grid is shown in Fig. 2. Daily LSE load profiles are adopted from the dynamic 5-bus example in [2]; see Fig. 3. Line capacities, reactance levels, and generator cost data are also adopted from [2].

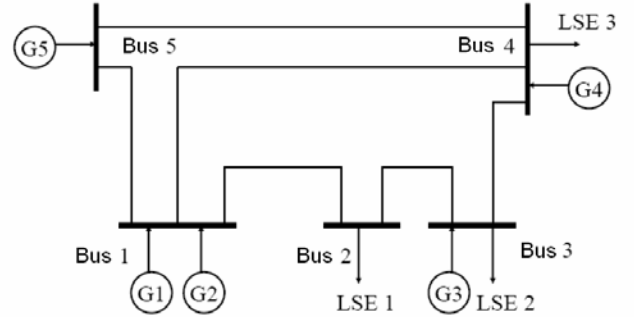


Fig. 2: 5-Bus Transmission Grid

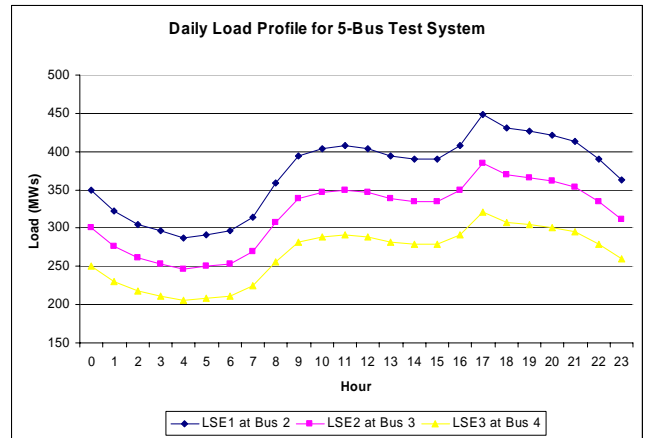


Fig. 3: 5-Bus Transmission Grid-Daily Load Profiles

Detailed solution values for the scenario in which suppliers submit their true production data to the MO ("the no-learning scenario") are given in [2].

This study simulates two Q-Learning scenarios for this 5-bus test case. In the first scenario the LSEs have relatively low critical values for curtailing demand, whereas in the second scenario they have relatively high critical values. Simulation results for these learning scenarios are compared with the no-learning scenario.

4.2 Review of Results from No-Learning Scenario

In the no-learning scenario analyzed in [2], each generator submits a supply offer that includes its true marginal cost function and its true generation upper limit. The MW production level of each generator and the LMP at each bus that are cleared by the MO based on true cost data from generators are depicted in Fig. 4(a) and Fig. 6(a).

Generators 3 and 5 are the two largest units in the system with a combined capacity 1120MW. The combined capacity of the three other small units is 410MW. The large units together with the high peak hour demand (1153.59MW) gives generators 3 and 5 potential market power. Note that the congestion between bus 1 and bus 2 exists for all 24 hours. This causes LMP separation between bus 1 and bus 2. During hour 17, the power flow on the line between buses 1 and 2 hits its upper thermal limit, and Generator 3 is dispatched at its upper production limit. Therefore, generator 4 that has the highest variable generation cost has to be dispatched to meet the demand. This results in a huge price spike at buses 2 and 3 at hour 17 that is about double of their LMP values at hour 16.

4.3 Results from the Two Learning Scenarios

Assume that the generators do not have to report their true marginal costs to the MO. Instead, the profit-seeking generators use Q-Learning to learn how to bid strategically to make more profits.

Since the system can be in several states, it does not have to stay in one single state in the long term. Rather, it may visit some states periodically or it may not even converge to a periodic pattern. Therefore, one has to define convergence in a different way. The Day-Ahead market is said to be convergent if, at any state, each generator chooses one action in that state with probability 0.99 or higher.

Due to the probabilistic nature of the learning algorithm, the simulation does not converge to the same values for each run. In order to average out the effects of different probabilities, 10 simulation runs are performed for each scenario and the mean values from the runs are reported.

In scenario one, LSEs have little tolerance for high LMPs. Their critical values for curtailing demand are only slightly higher than the LMPs that they will pay in the no-learning scenario. These critical values are given in Table 1.

Table 2 shows the number of days before convergence and final system states for 10 simulation runs. As can be seen, most of the time the system stays in state 8, in which every

LSE is curtailing demand every day. This implies that generators are using very aggressive bidding strategies, and making full use of their market power. In this case, generators actually are making more profits by moving the system to state 8 because, even in the situation of less demand in peak hour, the generators are still able to raise prices higher than the critical values of the LSEs.

Table 1: Critical Values for Learning Scenario 1 (\$/MWH)

| | LSE1 | LSE2 | LSE3 |
|----------------|-------|---------|------|
| Critical Value | 115.5 | 98.0251 | 47.5 |

Table 2: Number of Days before Convergence and Final States in Learning Scenario 1

| Simulation Run | 1 | 2 | 3 | 4 | 5 |
|------------------------------|----|-----|----|----|----|
| # of Days before Convergence | 83 | 189 | 81 | 77 | 77 |
| Final state(s) | 8 | 2↔8 | 8 | 8 | 8 |

| Simulation Run | 6 | 7 | 8 | 9 | 10 |
|------------------------------|-----|----|----|----|-----|
| # of Days before Convergence | 189 | 78 | 81 | 86 | 230 |
| Final state(s) | 2↔8 | 8 | 8 | 8 | 4↔8 |

2↔8 means the system visits state 2 and 8 periodically

In all 10 runs, all five generators converge by day 230 as shown in Table 2. The average number of days before convergence is 117.1. Note that in some cases the system moves back and forth between two states in a pattern of convergence.

In scenario two, the LSEs have high tolerance for high electricity prices. Their critical values for curtailing demand are higher than the critical values in scenario one. The critical values in this case are presented in Table 3. Table 4 shows the number of days before convergence and the final system states for the 10 simulation runs.

Table 3: Critical Values for Learning Scenario 2 (\$/MWH)

| | LSE1 | LSE2 | LSE3 |
|----------------|-------|-------|------|
| Critical Value | 135.5 | 115.5 | 55.5 |

Table 4: Number of Days before Convergence and Final States in Learning Scenario 2

| Simulation Run | 1 | 2 | 3 | 4 | 5 |
|------------------------------|-----|-----|-----|-----|-----|
| # of Days before Convergence | 222 | 325 | 259 | 177 | 236 |
| Final state | 1↔8 | 1 | 1↔8 | 2↔8 | 1↔8 |

| Simulation Run | 6 | 7 | 8 | 9 | 10 |
|------------------|-----|-----|-----|-----|-----|
| # of Days before | 222 | 252 | 227 | 226 | 241 |

| | | | | | |
|-------------|-----|-----|---|-----|-----|
| Convergence | | | | | |
| Final state | 1↔8 | 1↔8 | 1 | 1↔8 | 1↔8 |

In all 10 runs, all five generators converge by day 325. The average number of days before convergence is 238.7. Note that, most of the time, the system ends up visiting state 1 and state 8 in turn. The day of convergence comes later if the system keeps visiting more than one state. From the results it is seen that, in fact, Q-Learning allows the generators to take advantage of the LSEs, whose demand-side response only has one day memory. First, by submitting low supply offers, the generators make sure that the LSEs do not curtail their demand tomorrow. Afterward they submit high supply offers and profit significantly from the LSEs that decrease their peak hour demand tomorrow. Then the generators submit a low supply offer again and so on. The simulation results show that Q-Learning helps generators make more profits by sacrificing today's benefit for more profits in the future. This scenario is a good illustration of anticipatory reinforcement learning.

Differences between the learning scenarios and the no learning scenario are discussed below. Furthermore, it is desirable to know to what extent Q-Learning is capable of helping generators exercise market power. Fig. 4 depicts the mean values of MW production in learning scenarios 1 and 2, along with the corresponding simulation results obtained in the no-learning scenario. In the no-learning scenario, generator 4 is only dispatched at the peak hour. In both learning scenarios, in some simulation runs generator 4 is not dispatched. This is true when each generator is submitting an aggressive supply offer so that generator 4 is still the most expensive. However, in some simulation runs generator 4 chooses to submit less aggressive supply offers so that it becomes a relatively cheaper unit. Therefore, the average effect in the learning scenarios is that generator 4 is dispatched to some extent in each hour and has a steep increase in power output during the peak hour as it does in the no-learning scenario.

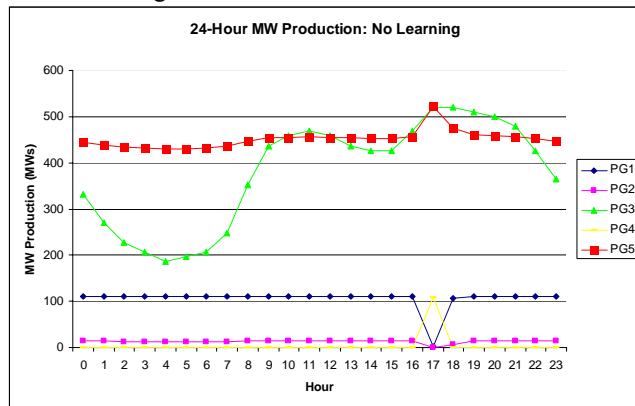


Fig 4.a: 5-Bus Transmission Grid Simulation Results for 24-Hour MW Production (No-Learning Scenario)

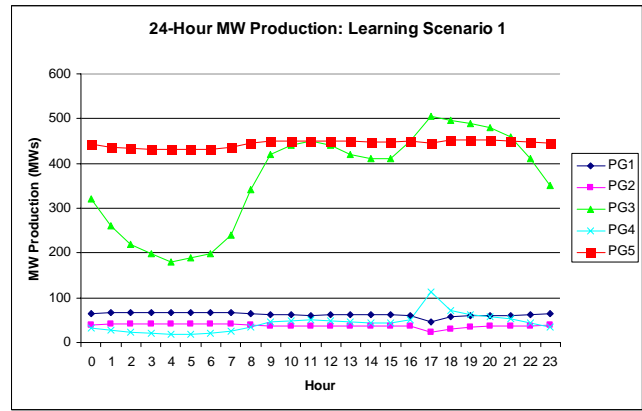


Fig. 4.b: 5-Bus Transmission Grid Simulation Results for 24-Hour MW Production (Learning Scenario 1)

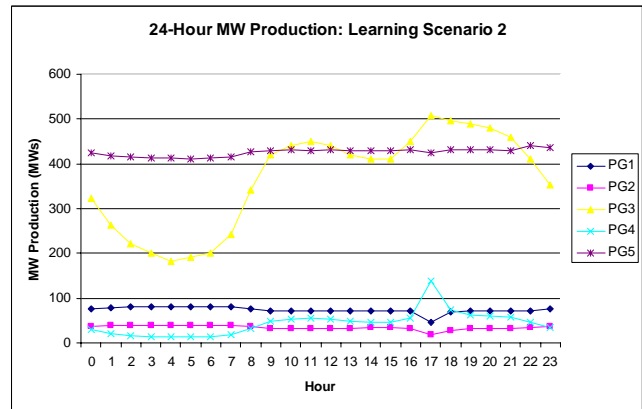


Fig 4.c: 5-Bus Transmission Grid Simulation Results for 24-Hour MW Production (Learning Scenario 2)

Figure 5 shows that the mean of the total profit gained by the generators in each learning scenario is much higher than what they made in the no-learning scenario. In fact, in the no-learning scenario the generators are not able to recover their fixed cost because they only covered their variable costs in their supply offers. This fact demonstrates that Q-Learning helps the generators to learn to exercise their potential market power to maximize their profits. Although not obvious, it can be observed in Fig 5 that, during peak hour 17, the generators are making more profits in learning scenario 2 than they are in learning scenario 1. The high level of tolerance for price spikes of the LSEs in learning scenario 2 gives the generators more opportunities to manipulate the market.

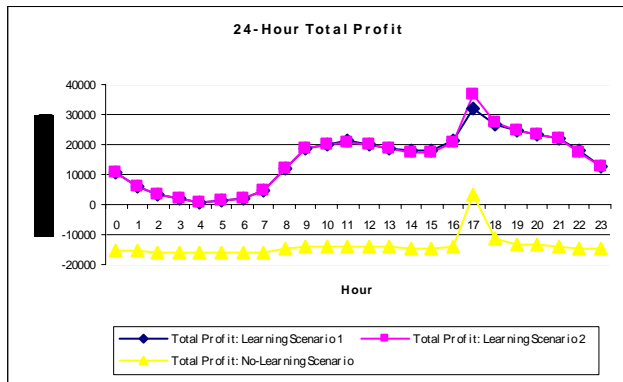


Fig. 5: 5-Bus Transmission Grid Simulation Results for 24-Hour Total Profits (No-Learning Compared with Learning Scenarios 1 and 2)

The 24-hour mean LMP values for the learning scenarios 1 and 2 along with the 24-hour LMP values for the no-learning scenario are shown in Fig. 6. In the no-learning scenario, the price spike at hour 17 is obvious. Although the LMPs in the learning scenarios 1 or 2 are substantially higher than for no-learning, the price fluctuation around the peak hour is much less. This finding is similar to the finding of Sun and Tesfatsion [2], who used reactive reinforcement learning to model the learning process of generators. However, since the sets of actions are different, one cannot draw a definitive conclusion about the learning techniques used in the two studies.

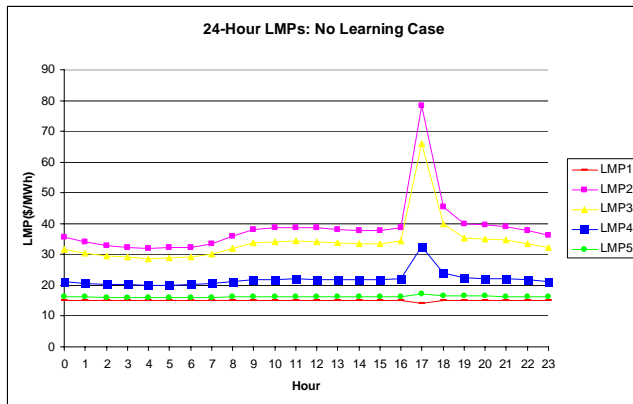


Fig. 6.a: 5-Bus Transmission Grid Simulation Results for 24-Hour LMPs (No-Learning Scenario)

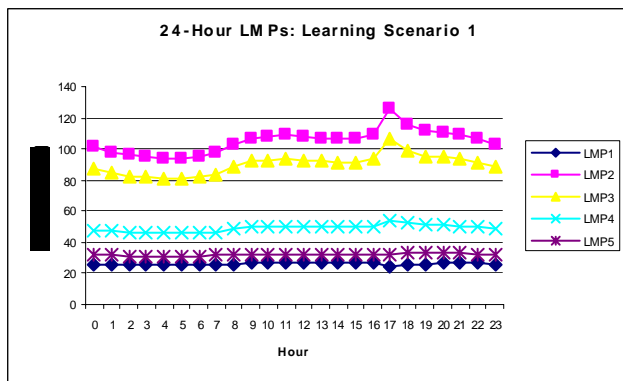


Fig. 6.b: 5-Bus Transmission Grid Simulation Results for 24-Hour LMPs (Learning Scenario 1)

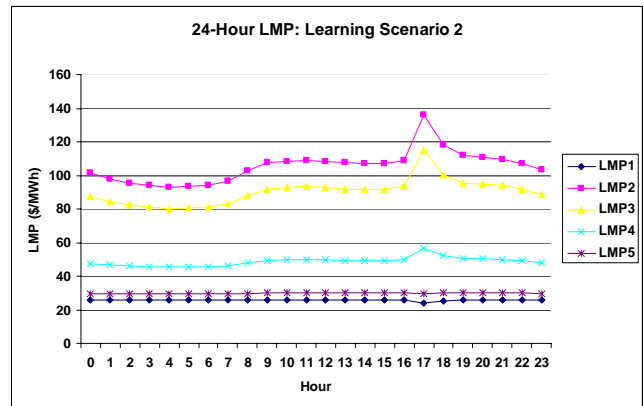


Fig. 6.c: 5-Bus Transmission Grid Simulation Results for 24-Hour LMPs (Learning Scenario 2)

5. Conclusion

This paper presents a novel application of Q-Learning to model electricity suppliers' learning behavior in a multi-agent electricity market environment. Demand-side response of LSEs is considered in the model, which provides a new way to define the system states of the model. Numerical simulations are performed on a 5-bus transmission grid test case to evaluate whether this implementation of Q-Learning is a good approach to modeling the learning behaviors of electricity suppliers.

Simulation results show that Q-Learning help electricity suppliers learn how to bid strategically under the condition of a simple demand-side response model. With Q-Learning capabilities, electricity suppliers find a way to make more profits in the long term by sacrificing their immediate profits.

Q-learning has some limitations. It assumes a finite domain of actions. Also, the Q-learning model developed in this research assumes that electricity suppliers do not explicitly take into account the presence of other electricity suppliers in the environments. Therefore the results may not be accurate when multiple agents interact in a realistic market environment. These limitations will be relaxed in future extensions of this research.

6. Acknowledgement

This research is sponsored by the Power System Engineering Research Center (PSERC) through a collaborative project involving Iowa State University, Washington State University and Smith College/Cornell University. The authors would like to thank Professors Gerald Sheblé, Anjan Bose and Judith Cardell and Mr. Jim Price, California ISO, for their useful discussions.

7. References

- [1] S. Borenstein, J. Bushnell, and F. A. Wolak, "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity

- Market,” Center for the Study of Energy Markets. Paper CSEMWP-102, June 2002.
- [2] J. Sun and L. Tesfatsion, “Dynamic Testing of Wholesale Power Market Designs: An Open-Source Agent-Based Framework,” ISU Economics Working Paper No. 06025, Revised April 2007. <http://www.econ.iastate.edu/tesfatsi/DynTestAMES.JSLT.pdf>
 - [3] V. Nanduri and T.K. Das, “A Reinforcement Learning Model to Assess Market Power under Auction-Based Energy Pricing,” *IEEE Transactions on Power Systems*, vol. 22, No.1, Feb. 2007.
 - [4] C. W. Richter, G. B. Sheblé, and D. Ashlock, “Comprehensive Bidding Strategies with Genetic Programming/Finite State Automata,” *IEEE Transactions on Power Systems*, vol. 14, pp. 1207-1212, Nov. 1999.
 - [5] J. Hu and M. P. Wellman, “Nash Q-learning for General-Sum Stochastic Games,” *J. Mach. Learn. Res.*, vol. 4, pp. 1039-1069, 2003.
 - [6] R. D. Zimmerman and D. Gan, “MATPOWER: A MATLAB Power System Simulation Package (Version 2.0),” Cornell University, New York 1997. <http://www.pserc.cornell.edu/>.
 - [7] C.J.C.H. Watkins, “Learning from Delayed Rewards,” Ph.D. Thesis, University of Cambridge, England, 1989.
 - [8] C.J.C.H. Watkins and P. Dayan, “Q-Learning,” *Machine Learning*, 3:279-292, 1992.

Nanpeng Yu received his B.Eng. from Tsinghua University, Beijing, China, in 2006. He is currently pursuing the Ph.D. degree at Iowa State University. His research interests are in strategic gaming in electricity markets and modeling of participants in power markets.

Chen-Ching Liu received his Ph.D. from the University of California at Berkeley. He is currently Palmer Chair Professor of Electrical and Computer Engineering at Iowa State University. Dr. Liu serves as President of the Council on Intelligent System Applications to Power Systems (ISAP). He is a Fellow of the IEEE.

Leigh Tesfatsion is Professor of Economics and Mathematics at Iowa State University. Her current research focuses on agent-based computational modeling, with a specific concern for market design issues in restructured electricity markets. She serves as Associate Editor for several economics and mathematics journals and is a Member of the IEEE.