Demand Response in Electricity Markets

Athina C. Tellidou, and Anastasios G. Bakirtzis, *Senior Member, IEEE*

Abstract-- **In this paper agent-based simulation is employed to study the effect of demand-side bidding in the exercise of monopoly power by generators. The energy market is formulated as a stochastic game, where each stage game corresponds to an hourly energy auction. Each hourly energy auction is cleared using Locational Marginal Pricing. Generators and consumers are modeled as adaptive agents capable of learning through the interaction with their environment, following a Reinforcement Learning algorithm. The SA-Q-learning algorithm, a modified version of the popular Q-Learning, is used. Test results on a two-node power system with two generator-agents and two consumer-agents, lead to some useful conclusions.**

***Index Terms*-- Electricity Markets, Reinforcement Learning, Demand-Side Bidding**

I. INTRODUCTION

MARKET power is one of the central topics in power market design and it has been studied extensively in the relative literature [1]. Market power is mainly exercised by generators, either by raising the price of their offer or by withholding capacity. Both these actions are difficult to be recognized as exercise of market power. Hence, one of the most challenging tasks in power market monitoring is to identify who exercises market power and how. The answers to these questions have important implications to power market design; therefore, many researchers have tried to approach this demanding issue in a satisfying way. Many suggestions have been made and the regulators have tried various schemes, in order limit these phenomena, but still no fulfilling answer has been given. Most of them agree in one thing: demand response may be the answer.

"Demand Response can be defined as the changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time", according to [1]. There is a wide range of categories of demand response programs. One of these categories is Market Based programs and the sub-category Demand Bidding [1] is the one studied in this paper.

Even though the lower price of electricity for consumers has been the primary concern, since the introduction of competition, consumers seem to have very little influence on the prices as well as on the design of the markets. The short-run price elasticity of the demand for electricity is small and this is due to the fact that, except for the large consumers, the rest have neither the financial incentive nor the expertise required to participate in such a complex and time-consuming task [2].

Little work has been done, according to the relative literature, on demand response and the reason is that it is still difficult to be fully implemented in a real market. In [1] the demand response is fully defined and the classification of demand response programs is presented. The benefits and the costs are outlined and the ways of demand response measurement are discussed. Tutorial paper [2] presents some characteristics of electricity markets from the demand-side point of view to highlight the importance of demand elasticity. It suggests that increasing the short-run price elasticity of the demand in electricity markets would improve their operation, but, since the price responsiveness of the demand is not in the best interest of generating companies, such market designs are unlikely to prevail without consumer pressure. In [3], a model for the electricity auctions is developed, which considers demand-side bidding. The consumers have the opportunity to submit bids for load reductions in specific periods of the 24-hour simulation that is presented. The results indicate the important effect of demand-side bidding in smoothing system marginal price and in mitigating price volatility.

In this paper, agent-based simulation is employed in order to study the impact of demand side bidding in an electricity market and, in particular, in the exercise of generators' market power. In Section II the electricity market structure is described and the problems faced by the Independent System Operator and the participants in the market, namely generators and consumers, are formulated. In Section III the basic elements of the static and repeated games are introduced, in order to model electricity market as a repeated game. In Section IV reinforcement learning theory and SA-Q-Learning algorithm are presented, the participants are modeled as agents who follow the specific algorithm and the agent-based market simulation is described. In Section V test results in a 2-node system include three examples, which indicate that demand-side bidding can indeed contribute to the limitation of power market exercise.

II. ELECTRICITY MARKET STRUCTURE

In this paper the behavior of generators and consumers in a spot market with hourly trading intervals is studied. Generators submit Energy Offers and consumers submit Energy Bids to the Independent System Operator (ISO). The Energy Offers declare the power generators are willing to sell at or above a certain price, while Energy Bids declare the power consumers wish to consume at or below a certain price. Each generator, $g \in \mathcal{G}$, with net capacity $P_g^{\max}$ and marginal cost mc_g, offers a certain amount of power P_g in MW,

Athina C. Tellidou and Anastasios G. Bakirtzis are with the Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, 54124, Thessaloniki, GREECE (e-mail: atellidou@eng.auth.gr, bakiana@eng.auth.gr).

$0 \le P_g \le P_g^{\max}$, at a constant price b_g in €/MWh, which cannot be higher than the market price cap, pc , or lower than the generator's marginal cost, $mc_g \le b_g \le pc$. Similarly, each consumer, $f \in \mathcal{F}$, with maximum load demand $D_f^{\max}$, bids a certain amount of power D_f in MW, $0 \le D_f \le D_f^{\max}$, at a constant price b_f in €/MWh, which cannot be higher than the marginal utility, mu_f , $b_f \le mu_f$

A. The ISO Market Clearing Problem

In this paper the symmetric electricity market is simulated, where the players – both generators and consumers – submit strategically their offers and bids. The ISO collects the energy offers (P_g, b_g) of all generators $g \in \mathcal{G}$, and the energy bids (D_f, b_f) of all consumers $f \in \mathcal{F}$, and computes the quantities $p_g, \forall g \in \mathcal{G}$ and $d_f, \forall f \in \mathcal{F}$, as well as the nodal prices, $LMP_k, \forall k \in \mathcal{K}$, that clear the market, by solving the following optimal power flow (OPF) problem:

$$Max \sum_{f \in \mathcal{F}} b_f \cdot d_f - \sum_{g \in \mathcal{G}} b_g \cdot p_g \tag{1}$$

Subject to:

$$\mathbf{B} \cdot \boldsymbol{\theta} = \mathbf{H}_{\mathcal{KG}} \mathbf{p} - \mathbf{H}_{\mathcal{KF}} \mathbf{d} \tag{2}$$

$$\theta_{ref} = 0 \tag{3}$$

$$\left| \frac{1}{x_{km}} (\theta_k - \theta_m) \right| \le F_{km}^{max} \quad \text{for all lines } km \tag{4}$$

$$0 \le p_g \le P_g \quad \text{for all generators } g \in \mathcal{G} \tag{5}$$

$$0 \le d_g \le D_g \quad \text{for all consumers } f \in \mathcal{F} \tag{6}$$

where:

$\mathbf{p}$	generator active power output vector
$\mathbf{d}$	consumer active power demand vector
$\boldsymbol{\theta}$	bus voltage phase angle vector
θ_{ref}	reference bus voltage phase angle
x_{km}	reactance of line km
$F_{km}^{\max}$	transmission capacity limit of line km
$\mathbf{B}$	network admittance matrix
$\mathbf{H}_{\mathcal{KG}}$	bus to generator incidence matrix (size $\mathrm{K} \cdot \mathrm{G}$)
$\mathbf{H}_{\mathcal{KF}}$	bus to consumer incidence matrix (size $\mathrm{K} \cdot \mathrm{F}$)

Constraints (2) represent the system DC power flow equations, (3) defines the slack bus voltage phase angle since $\det(\mathbf{B}) = 0$ [1], (4) represent the transmission line power flow limits, and (5) and (6) represent the unit active power output limits and the maximum load consumption, respectively. The Lagrange multipliers of the nodal active power balance constraints (2) are the nodal prices (LMPs).

[1] The slack bus is not omitted in (2).

B. The Generator Profit Maximization Problem

The Generator's objective is to maximize his profits in the spot market, by selecting the parameters of his energy offer, (P_g, b_g) . Hence, his optimization problem is formulated as follows:

$$Max\ Profit_g = (LMP_{k(g)} - mc_g) \cdot p_g \tag{7}$$

Subject to the following constraints:

$$0 \le P_g \le P_g^{\max} \tag{8}$$

$$mc_g \le b_g \le pc \tag{9}$$

as well as the ISO market clearing problem solution (1)-(6), needed to define p_g and $LMP_{k(g)}$ used in (7).

However, the generator does not have the information on the transmission network, the consumer energy bids and the competitor energy offers that the ISO has when solving the market-clearing problem. Section IV describes a reinforcement learning process by which the generator can "learn" through the repetition of the hourly energy auction to select the profit maximizing parameters of his energy offer, (P_g, b_g), based only on publicly available information (LMPs and previous energy offers).

C. The Consumer Net Utility Maximization Problem

Consumer's objective is to maximize his net utility in the spot market, by selecting the parameters of his energy bid, (D_f, b_f) . Hence, his optimization problem is formulated as follows:

$$Max\ NetUtility_f = (mu_f - LMP_{k(f)}) \cdot d_f \tag{10}$$

Subject to the following constraints:

$$0 \le D_f \le D_f^{\max} \tag{11}$$

$$0 \le b_f \le mu_f \tag{12}$$

as well as the ISO market clearing problem solution (1)-(6), needed to define d_f and $LMP_{k(f)}$ used in (10).

However, the consumer does not have the information on the transmission network, the generators energy offers and the competitors energy bids that the ISO has when solving the market-clearing problem. Section IV describes a reinforcement learning process by which the consumer can "learn" through the repetition of the hourly energy auction to select the profit maximizing parameters of his energy bid,

$\left(D_f, b_f\right)$, based only on publicly available information (LMPs and previous energy bids).

III. Repeated Games

A. Static Games

Let us assume the following n-player simple game:

Players 1 through n simultaneously choose actions $a_1 \in \mathcal{A}_1$ through $a_n \in \mathcal{A}_n$, respectively.

They receive their payoffs $u_1(a_1,\ldots,a_n)$ through $u_n(a_1,\ldots,a_n)$.

$\Gamma = \left\{\mathcal{A}_1,\ldots,\mathcal{A}_n; u_1,\ldots,u_n\right\}$ denotes the above static game [4].

According to the above, the market operation, as described in section 2, can be seen as a stage game, where $\mathcal{G} \cup \mathcal{F}$ is the set of players; then each generator and each consumer can be regarded as players and their action space consists of all the possible selections of their energy offer, $\left(P_g, b_g\right)$, and energy bid, $\left(D_f, b_f\right)$, respectively.

B. Infinitely Repeated Games

Definition 1: Given a stage game Γ, let $\Gamma(\infty,\delta)$ denote the infinitely repeated game in which Γ is repeated forever and the players share the discount factor δ. For each t, the outcomes of the $t-1$ preceding plays of the stage game are observed before the t^{th} stage begins. Each player's payoff in $\Gamma(\infty,\delta)$ is the present value of the player's payoffs from the infinite sequence of stage games [4].

Definition 2: Given the discount factor δ, the present value of the infinite sequence of payoffs $u^{(1)}, u^{(2)}, u^{(3)}, \ldots$ is:

$$u^{(1)} + \delta u^{(2)} + \delta^2 u^{(3)} + \cdots = \sum_{t=1}^{\infty} \delta^{t-1} u^{(t)} \qquad (13)$$

The discount factor reflects the time value of money [4]; the closer δ is to 1 the more important are distant payoffs. In games of complete information, the discount factor affects the outcome of the repeated game. The analysis presented in this paper – which refers to a game of incomplete information – is indifferent about the value of the discount factor.

According to Definition 1, the electricity market simulated in this paper can be thought of as an infinitely repeated game, where each stage game is the same with the one described in the last paragraph of the previous subsection.

IV. Player's Behavior Under Modified Q-Learning

A. Reinforcement Learning

Many learning theories developed as a result of man's effort to analyze the behavior of animals and artificial systems. RL is one of them and focuses on the effect of rewards (positive payoffs) and punishments (negative payoffs) on subjects' choices in their attempt to achieve a goal [5].

RL theory's basic elements are:

- the learner or the decision-maker, who is called the *agent*, and
- everything it interacts with, which is called the *environment.*

The effects of actions cannot be fully predicted; thus the agent must monitor its environment frequently and react appropriately. As it becomes clear, the basic concept behind RL is *trial and error* search, since the agent explores its environment and learns from his mistakes [5].

B. Q-Learning Algorithm

Q-Learning algorithm, proposed by Watkins [6], is one of the most commonly used RL algorithms, because of its simplicity. Its main advantages are that it can be used on-line and it is model free – it does not need an explicit model of its environment. Q-Learning is an algorithm for learning to evaluate the payoff for a given state-action pair. In order for the algorithm to be suitable for our game-theoretic multi-agent approach, some modifications – as presented in [7] – of the original algorithm, concerning the Q values, have been adopted.

The agent n – the agent could be either a generator, g, or a consumer, f – in Q-Learning keeps in memory a function $Q_n(a_n)$ [2], that represents the expected payoff he believes he will obtain by taking an action a_n; the function of the expected payoff is represented by an one-dimensional lookup table indexed by actions, whose elements are defined as Q-values.

Let $\mathcal{A}_n = \left\{a_{n,1}, a_{n,2}, \ldots, a_{n,A_n}\right\}$ be the set of A_n possible actions the agent n can take. In the t^{th} stage, the agent:

1. Selects and performs an action $a_n^{(t)} \in \mathcal{A}_n$ using a policy.
2. Receives an immediate payoff $u_n^{(t)}(a_n)$.
3. Updates his Q values according to:

$$Q_n^{(t)}(a_n) = \begin{cases} (1-\alpha_n^{(t)}(a_n))Q_n^{(t-1)}(a_n) + \alpha_n^{(t)}(a_n)u_n^{(t)}(a_n) & \text{if } a_n = a_n^{(t)}, \\ Q_n^{(t-1)}(a_n) & \text{otherwise} \end{cases} \qquad (14)$$

According to equation (14), only Q values corresponding to the last action chosen are updated. $\alpha_n^{(t)}(a_n)$ is a learning rate in the range [0,1), that reflects the degree to which estimated Q values are updated by new data; it can be different in each episode and action dependent [8].

C. SA-Q-Learning Algorithm

Use The SA-Q-Learning algorithm was proposed by Guo, Liu and Malec [9] as a result of their research in controlling the balance between exploration and exploitation during the evolution of the Q-learning algorithm. They applied the Metropolis criterion [10], used in the Simulated Annealing (SA) algorithm [11], in order to determine the action-selection

[2] The Q-values of each agent depend on the actions selected by other agents also; however, for simplification reasons, we use this notation.

strategy of Q-learning. The outcome is very promising, as shown in their experiments, since in the execution process of the Q-learning algorithm the exploration gradually decays, leading to convergence.

The SA-Q-Learning algorithm explains the first step of the Q-learning algorithm, by defining the followed policy, i.e. the criteria an agent uses to select the next action. Hence, the SA-Q-learning can be described by the steps 1-3 mentioned in the previous subsection, with the first step being replaced by the following actions:

a. Selects an action $a_{n,r} \in \mathcal{A}_n$ randomly.

b. Selects an action $a_{n,p} \in \mathcal{A}_n$ following a greedy[3] policy: $a_{n,p} = \arg\max_{a_n} Q_n^{(t-1)}(a_n)$.

c. Generates a random number $\xi \in (0,1)$.

d. Selects and performs action $a_n^{(t)} \in \mathcal{A}_n$ as follows:

$$a_n^{(t)} = \begin{cases} a_{n,p} \ \ if \ \xi \geq \exp\left[\dfrac{Q_n^{(t-1)}(a_{n,r}) - Q_n^{(t-1)}(a_{n,p})}{T_n^{(t)}}\right] \\ a_{n,r} \ \ otherwise \end{cases} \tag{15}$$

e. Calculates $T_n^{(t+1)}$ by the temperature-dropping criterion.

Although the temperature-dropping criterion can be in general arbitrary, in this paper the geometric scaling factor criterion is used, as in [9]. Let $T_n^{(t)}$ be the Temperature in the t^{th} stage and $\lambda \in (0.5,1)$ a constant, usually close to 1, in order to guarantee a slow decay of the temperature in the algorithm. Then in the $t+1$ stage the temperature will be:

$$T_n^{(t+1)} = \lambda T_n^{(t)} \ , t = 0,1,2,\ldots \tag{16}$$

D. Modeling Player's Behavior

Each player must select his actions in every stage in order to maximize his payoff. The application of the SA-Q-learning algorithm in modeling the player behavior requires the definition of the admissible *actions* and the returned *payoff*.

Action. The generator-agent action is the selection of the offer quantity and price, $\left(P_g, b_g\right)$, while the consumer-agent action is the selection of the bidding quantity and price, $\left(D_f, b_f\right)$. The generator's action space is discretized, by discretizing both the offer quantity and the offer price intervals of variation, (8) and (9), into A_g^P and A_g^b levels respectively. Similarly, the consumer's action space is discretized, by discretizing both the bidding quantity and the bidding price intervals of variation, (11) and (12), into A_f^D and A_f^b levels respectively.

Payoff. The payoff received by each generator during an auction round is equal to the profit, in €, the agent makes by participating in the spot market, defined in (7). The payoff received by each consumer during an auction round is equal to the net utility, in €, the agent makes by participating in the spot market, defined in (10).

The learning rate is designed to be action dependent, as in [8]. The learning rate $\alpha_n^{(t)}(a_n)$ is inversely proportional to the visited number $\beta_n^{(t)}(a_n)$ of action a_n up to the present trading stage, as follows:

$$\alpha_n^{(t)}(a_n) = \frac{1}{\beta_n^{(t)}(a_n)} \tag{17}$$

E. Agent Based Energy Market Simulation

The spot energy market simulation consists of the repetition for a large number of stages, $t = 0,1,2,\ldots,t^{\max}$ of the following steps:

Step 1: All agents select an action $a_n^{(t)} \equiv \left(P_n^{(t)}, b_n^{(t)}\right)$ according to the policy defined by the SA-Q Learning (15), and submit their energy offers and energy bids defined by the selected action to the ISO.

Step 2: The ISO processes the offers and the bids submitted by all agents, along with transmission system, and computes the quantities and prices that clear the market by solving (1) -(6).

Step 3: The ISO posts the public information on nodal prices, $LMP_k, \forall k \in \mathcal{K}$, and informs every generator $g \in \mathcal{G}$ about the quantity, p_g, of his energy offer accepted in the spot market and every consumer $f \in \mathcal{F}$ about the quantity, d_f, of his energy bid accepted in the spot market.

Step 4: All agents use the information they receive from the ISO (Step 3) to compute profits and net utilities and update their Q Tables according to (14).

Step 5: The stage count, t is updated; the Temperature is updated according to (16); the learning rate is updated according to (17). The whole process, Step 1 through Step 5, is repeated if the stage count, t, is less than the maximum number of stages, $t^{\max}$.

V. Test Cases

A simple, two-node system, shown in Fig. 1 is used in our test cases. The transmission capacity limit is 100 MW. The generator data are shown in Table I. The marginal utility for both consumers is set equal to 40 €/MWh. Locational Marginal Pricing is used for market settlement, as already discussed. The market price cap is 40 €/MWh.

Three cases are examined:

In Case A each generator offers its full capacity at marginal cost, so that competitive prices result. This case is used as reference to test the exercise of market power by the generators.

[3] Greedy policy: the agent always selects the action with the highest Q value.

Fig. 1. Two-Node Test System

TABLE I
GENERATION DATA.

Nodes k	Generators g	P_g^{max} [MW]	mc_g [€/MWh]
1	Gen-1	300	15
2	Gen-2	300	30

In Case B each generator participates in a repeated energy auction trying to maximize its profits, by reinforcement learning, as described in Section IV. This case is used as reference to test the effect of demand-side bidding.

In Case C both generators and consumers participate in a repeated auction trying to maximize their profits and net utilities, respectively, by reinforcement learning, as described in Section IV.

Parameter Selection. All generators and consumers are considered to be players in our market and their behavior is modeled through the SA-Q-learning algorithm, as described before. The parameters of the algorithm that need to be defined are the initial temperature $T^{(0)}$ and the constant λ of the temperature-dropping criterion. All generators have the same parameters $T^{(0)} = 100{,}000$ and $\lambda = 0.99$.

A. Reference Case

Both generators offer their full capacity at marginal cost, so that competitive prices result. The market clearing results under competitive prices are presented in Table II. Owing to the 100 MW transmission limit, the cheaper generator, Gen-1, is dispatched only up to 200 MW. The remaining 100 MW of the Node-2 demand are supplied by the more expensive local generator, Gen-2. There is Locational price difference, owing to congestion, and the LMP at each node is equal to the marginal cost of the local generator. Hence, neither generator makes profit from the energy market.

TABLE II
MARKET CLEARING UNDER COMPETITIVE PRICES (REFERENCE CASE)

Nodes k	Generators g	P_g^{disp} [MW]	LMP_k [€/MWh]	$Profit_g$ [€]
1	Gen-1	200	15	0
2	Gen-2	100	30	0

TABLE III
GENERATOR OFFERS UNDER REINFORCEMENT LEARNING (DUOPOLY CASE)

Nodes k	Generators g	P_g [MW]	b_g [€/MWh]
1	Gen-1	200	25
2	Gen-2	>100	40

TABLE IV
MARKET CLEARING UNDER REINFORCEMENT LEARNING (DUOPOLY CASE)

Nodes k	Generators g	P_g^{disp} [MW]	LMP_k [€/MWh]	$Profit_g$ [€]
1	Gen-1	200	40	5000
2	Gen-2	100	40	1000

B. Duopoly Case

When both generator-agents act strategically, trying to maximize profits through reinforcement learning, the resulting market conditions are shown in Tables III and IV. As shown in Table III the first agent identifies his potential market power and exercises it by withholding output. He offers only 200 MW, in order to leave the transmission line un-congested and be paid at the LMP of Node-2. Hence, he manages to raise the LMP of Node-1 to a level certainly greater than 30 €/MWh and increase his profits, from 0 € to a level grater than (30 – 15) * 200 = 3000 €. Despite the fact that Gen-1 does not know the transmission capacity, he "learns" to withhold capacity through repetition (in [1] a deeper analysis of this case is presented).

Since there is no congestion, both producers are paid at the same system-wide Market Clearing Price, MCP. The resulting MCP is equal to the market price cap, as shown in Table IV owing to the fact that agent Gen-2 realizes his monopoly power of the local, Node-2, demand and the absence of demand elasticity at Node-2. Hence, he chooses to exercise his monopoly power and offer his quantity to the higher price he can; consequently, his offer price is equal to the market price cap, as shown in Table III, and his profits has been increased from 0 € to 1000 €.

C. Demand-Side Bidding Case

When both generators and consumers act strategically, trying to maximize profits through reinforcement learning, the resulting market conditions are shown in Tables V-VIII. As shown in Table VI the consumers choose to decrease their aggregate bidding quantity down to the limit of 200 MW. This is the quantity that can be fully covered by the first – cheaper – generator; hence the second generator cannot exercise the monopoly power, he has due to the network topology. Gen-1 realizes that he is the only supplier and, as shown in Table V, he offers more than 200 MW, increasing his offer price up to the level of Gen-2's marginal cost, at 30 €/MWh. Hence, through the learning procedure, Gen-1 recognizes his monopolistic position and take advantage of it up to the point Gen-2 is not entering into the competition. Observing Table

VII, where the results of the clearing of the market for generators are displayed, we conclude that the profits are considerably decreased compared to the case of duopoly. Hence, the demand-side bidding managed to limit the generators exercise of market power and lower the MCP.

TABLE V
GENERATOR OFFERS UNDER REINFORCEMENT LEARNING
(DEMAND-SIDE BIDDING CASE)

Nodes k	Generators g	P_g [MW]	b_g [€/MWh]
1	Gen-1	250	30
2	Gen-2	>50	>30

TABLE VI
CONSUMER BIDS UNDER REINFORCEMENT LEARNING
(DEMAND-SIDE BIDDING CASE)

Nodes k	Consumers f	D_f [MW]	b_f [€/MWh]
1	Con-1	100	>30
2	Con-2	100	30

TABLE VII
MARKET CLEARING FOR GENERATORS UNDER REINFORCEMENT LEARNING
(DEMAND-SIDE BIDDING CASE)

Nodes k	Generators g	P_g^{disp} [MW]	LMP_k [€/MWh]	$Profit_g$ [€]
1	Gen-1	200	30	3000
2	Gen-2	0	30	0

TABLE VIII
MARKET CLEARING FOR CONSUMERS UNDER REINFORCEMENT LEARNING
(DEMAND-SIDE BIDDING CASE)

Nodes k	Consumers f	D_f^{disp} [MW]	LMP_k [€/MWh]	$NetUtility_f$ [€]
1	Con-1	100	30	1000
2	Con-2	100	30	1000

D. Simulation Environment

The agent based simulation has been developed in JBuilder 2005 environment using Java (J2SE 5.0). A commercial package, GAMS 2.5 (CPLEX solver), is used for the solution of the ISO market clearing problem. All simulations run on an AMD Athlon ™ Processor 3200+, 2.01 GHz, 1.75 GB RAM.

VI. CONCLUSIONS

An analysis of demand-side bidding in a simulated electricity market was presented in this paper. The electricity market was formulated as a repeated game, where each hourly auction is represented by a stage of the game. For the needs of the analysis agent-based simulation was employed, where each generator and each consumer was modeled as adaptive agents, following a SA-Q-learning bidding behavior.

Test cases on a simple two-node test system, with two generators and two consumers led to the following conclusions. Consumers can learn to recognize the exercise of market power from generators and trough their bidding behavior can limit it. Hence, even when their behavior is simulated through the simple model presented in Section IV, they can contribute to the decrease of electricity prices.

VII. REFERENCES

[1] A. Tellidou, A. Bakirtzis, "Agent-Based Analysis of Capacity Withholding and Tacit Collusion in Electricity Markets," *IEEE Transactions on Power Systems*, Vol. 22, No. 4, pp. 1735-1742, Nov. 2007.

[2] M. H. Albadi, E. F. El-Saadany, "Demand Response in Electricity Markets: An Overviw," in *Proc. 2007 IEEE Power Engineering Society General Meeting,* 24-28 June 2007.

[3] D. S. Kirschen, "Demand-Side View of Electricity Markets," *IEEE Transactions on Power Systems*, Vol. 18, No. 2, pp. 520- 527, May 2003.

[4] A. Borghetti, G. Gross, C. A. Nucci, "Auctions with Explicit Demand-Side Bidding in Competitive Electricity Markets", Chapter 4 of "The Next Generation of Unit Commitment Models", B.F. Hobbs, M. Rothkopf, R.P. O'Neill, H.P. Chao, Kluwer Academic Press, Boston, 2001.

[5] R. Gibbons, "A Primer in Game Theory," Harvester Wheatsheaf, 1992.

[6] R. S. Sutton, A. G. Barto, "Reinforcement Learning: An Introduction," MIT Press, Cambridge, MA, 1998

[7] C. J. C. H. Watkins, P. Dayan, "Technical Note: Q-Learning," Machine Learning, 8, pp. 279-292, (1992)

[8] T. Krause, G. Andersson, D. Ernst, E. V. Beck, R. Cherkaoui, A. Germond, "A Comparison of Nash Equilibria Analysis and Agent-Based Modeling for Power Markets," *Proceedings of the 15th PSCC*, Liege, Belgium, August 22-26, 2005.

[9] G. Xiong, S. Okuma, H. Fujita, "Multi-Agent Experiments on Uniform Price and Pay-as-Bid Electricity Auction Markets", *Proceedings of the International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004),* 2004 IEEE, April 2004, Hong Kong.

[10] M. Guo, Y. Liu, J. Malec, "A new Q-Learning Algorithm based on the Metropolis Criterion," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, Vol. 34, No. 5, pp. 2140- 2143, Oct. 2004.

[11] N. Metropolis, A. W. Rosenbluth, M. N. Rodenbluth, A. H. Teller, E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, Vol. 21, No. 6, pp 1087-1092, June 1953.

[12] S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, Vol. 220, No. 4598, pp 671-680, May 1983.