Paper

A Supplier Bidding Strategy Through Q-Learning Algorithm in Electricity Auction Markets

Gaofeng Xiong* Student Member
Tomonori Hashiyama** Member
Shigeru Okuma* Member

One of the most important issues for power suppliers in the deregulated electric industry is how to bid into the electricity auction market to satisfy their profit-maximizing goals. Based on the Q-Learning algorithm, this paper presents a novel supplier bidding strategy to maximize supplier's profit in the long run. In this approach, the supplier bidding strategy is viewed as a kind of stochastic optimal control problem and each supplier can learn from experience. A competitive day-ahead electricity auction market with hourly bids is assumed here, where no supplier possesses the market power. The dynamics and the incomplete information of the market are considered. The impacts of suppliers' strategic bidding on the market price are analyzed under uniform pricing rule and discriminatory pricing rule. Agent-based simulations are presented. The simulation results show the feasibility of the proposed bidding strategy.

Keywords: Deregulation, day-ahead electricity auction market, Q-Learning algorithm, supplier bidding strategy.

1. Introduction

In the past decade, the electric utility industry in many countries around the world has been undergoing fundamental structural changes to introduce competition and enhance efficiency. The traditional vertically integrated utility is deregulated to open up the system to the market, in response to the pressures of privatization and customer demands. Electricity and services can be sold and purchased as a commodity through different market structures. Under this deregulated and competitive environment, economics and profitability have become the major concern of every market participant, and each of them will act in his/her own self-interest in this new environment.

Among the proposed market structures, the electricity auction market has been widely experienced and implemented in different countries with different protocols. Market participants – electricity suppliers, and distribution companies – are required to submit their sealed bids to the auction market to compete for power energy. All participants winning the auction will be paid based on the rules agreed upon by the participants. Thus, the bidding strategy, which is essential for a successful business in this auction market, is becoming one of the most important issues in deregulated electric industry. Market participants can greatly improve their benefits by strategic bidding.

* Okuma Lab., Dept. of Information Electronics, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
** Hashiyama Lab., Institute of Natural Sciences, Nagoya City University,
Mizuho-cho, Mizuho-ku, Nagoya 467-8501, Japan

On the other hand, the current electricity market models, which are being implemented in many countries, are far from perfect and mature, and therefore, are still in the evolving process. Many issues that are being raised need to be solved, such as issues concerning market designing, system security and congestion management. California's electricity crisis has presented us these complex issues and challenges. One of lessons from California's electricity crisis is that market design and rules should assure markets workably competitive, and making comprehensive simulations and studies is necessary before critical decisions are made. It is believed that, developing profitable strategies for individual firms under different market designs can provide a deep insight into the complex new electricity markets and identify how rules can be altered to improve the performance of the market [1].

Developing bidding strategies for competitive suppliers has been studied by many researchers in recent years. Game theory [2] is naturally the first choice to deal with this issue and lots of works have been done using this traditional theory. In Ref. (3), a Nash game approach is used to study the pricing strategy in the deregulated power marketplace, where each participant has incomplete information about others. A method using Cournot non-cooperative game theory to determine the optimal supply quantity for each power producer in an oligopoly electricity market is presented in Ref. (4). The results show that the estimation accuracy of production cost functions of rivals plays an important role in this market. Different electricity market rules and their effects on bidding behaviors in a non-congestion grid are analyzed in Ref. (5). The authors conclude that generators can take advantage of congestion in their strategic

bidding behavior.

But game theory is not the only solution to this problem. In fact, due to the complexity, dynamics and uncertainty of the restructured electricity market, evolutionary computation algorithms and reinforcement learning are receiving increasing attention recently and becoming major tools in solving this problem. A genetic algorithm is developed in Ref. (6) to evolve the bidding strategies of participants in a double auction market. Markov Decision Process is used to optimize the bidding decisions to maximize the expected reward over a planning horizon in Ref. (7). The optimal bidding problem is modeled as a stochastic optimization problem in Ref. (8), and, a Monte Carlo approach based method and an optimization based method are developed to solve this problem. An agent-based simulation model of a wholesale electric market is developed in Ref. (15) to provide a source for strategic insight into the diverse aspects of the emerging electricity marketplace, and Q-Learning algorithm is used to generate the price offers for generation companies in a bilateral contracts market for electricity.

We assume that with the development and wide use of new power generation technology, the number of Independent Power Producers (IPPs) requesting interconnection to various locations has largely increased. These IPPs participate in a day-ahead electricity auction market, competing against each other to supply electricity. The bidding strategy for these IPPs is viewed as one kind of stochastic optimal control problem known as the Markovian Decision Problem (MDP) [18], and Q-Learning algorithm [18]~[20] is used to develop an optimal bidding strategy for suppliers to satisfy their profit-maximizing goals in a long term, rather than in a particular round of auction.

It is assumed that no supplier possesses the market power, which can be used to manipulate the market price to satisfy his/her own interest. Each market participant in this market is assumed to have only information on his/her own cost and the publicly available information of the market, but lack information on other participants. The market participants are also assumed to be so many that it is very difficult for each supplier to estimate other participants' bidding behaviors. But, each participant is designed to have the ability to use the public information of the market and to learn from experience,

In this paper, we also attempt to provide some views on two controversial issues in the evolving electricity auction market:

- *Which market pricing rule is better for electricity auction market, the uniform pricing rule or the discriminatory pricing rule?* Although the uniform pricing rule is theoretically superior to the discriminatory pricing rule and the majority of pricing rule adopted in electricity auction markets now is the uniform pricing rule, this issue is still an open question. In fact, the new electricity trading arrangement (NETA) of U.K. electricity market [9] has replaced the mandatory daily uniform price auction with a discriminatory price auction (a switch from the uniform to discriminatory pricing rule).
- *Should bidders with the lowest costs of production bid their true costs under the uniform pricing rule?* It is widely believed that, if all generators were to be paid on market clearing price under the uniform pricing rule, the dominant bidding strategy of a bidder, especially of bidder with the lowest costs of production, is to bid the true cost. But, the author concludes in Ref. (10) that, bidders have no incentives to bid their true costs under the uniform pricing rule; instead, they will likely mask their bids above their costs of production.

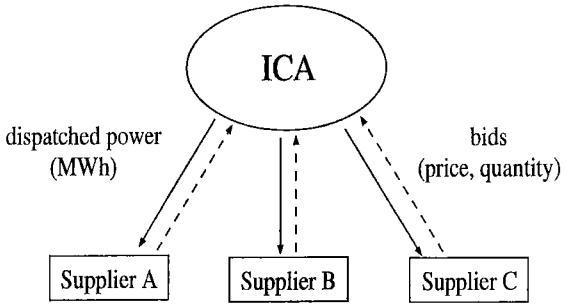


Fig. 1. The relationship of the ICA and suppliers.

This paper is organized as follows: Section 2 describes the model of a day-ahead electricity auction market. Section 3 presents the Q-Learning algorithm and the proposed supplier bidding strategy. Section 4 shows the simulation results, which is based on a multi-agent simulation approach. Section 5 gives the conclusions.

2. A Day-ahead Electricity Auction Market

A day-ahead electricity auction market with no demand-side bidding is assumed here. In this day-ahead auction market, all suppliers wishing to sell power tomorrow must submit their bids today to an Independent Contract Administrator (ICA) [17], who will clear the market, determine which supplier should be used to meet the forecasted load, and check if the security and reliability constraints of the power system are satisfied. The relationship of the ICA and suppliers is shown in Fig.1.

Everyday, suppliers submit their 24 separate hourly bids with price ($/MWh) and quantity (MW), at which they are willing to sell, to compete for the power loads over the 24-hour of the next day, which are forecasted by the ICA. The bids from suppliers are ranked by the ICA from the cheapest to the most expensive to construct a supply curve on an hourly basis. The ICA will then select the cheapest supplier until the load of each hour of the next day is met.

At the end of every trading day, each supplier is notified of his hourly dispatched power (MWh), which is the quantity called into operation during the next day, and the hourly market price ($/MWh), which is assumed to be the only publicly available information to each supplier in this paper.

The market winners are paid based on the rules agreed by all market participants. Currently, there are two ma-

jor pricing rules used in the deregulated electricity auction markets around the world.

- Uniform pricing rule: all winners are paid at the market clearing price (MCP), which is the highest bid price of winners ("pay marginal").
- Discriminatory pricing rule: each winner is paid at his own bid price ("pay as bid").

It is assumed that the publicly available hourly market price is the hourly market clearing price under uniform price rule, or the hourly sales-weighted average bid prices of all suppliers under discriminatory pricing rule.

In general, increasing the amount of information available to all bidders could increase the efficiency of the auction market. Therefore, it is better for the Independent Contract Administration (ICA) to publish other information, such as the maximum and minimum bid prices of everyday, market power demand and weather conditions, as well as the hourly market price. But, the problem is that increasing the amount of publicly available information could at the same time lead to the risk of making the unexpectedly collusive behavior between bidders easier to implement. Therefore, we assume in this paper that the hourly market price is the only publicly available information to all bidders in an attempt to reduce the potentiality of collusive behavior between bidders.

3. Developing Bidding Strategy Through Q-Learning Algorithm

Q-Learning (QL) algorithm is a reinforcement learning algorithm [18] proposed by Watkins for solving the Markovian Decision Problems with incomplete information. It does not need an explicit model of its environment and can be used on-line to find the optimal strategy through experience obtained from the direct interaction with its environment. These features make it well suitable for dealing with the decision-making problems in the repeated games against unknown opponents, such as the bidding strategy in the electricity auction market. Based on the Q-Learning algorithm, this section provides an optimal bidding strategy for suppliers to maximize their profits in the day-ahead electricity auction market.

3.1 Q-Learning Algorithm Assume that a learning agent interacts with its environment at each of a sequence of discrete time steps, $t = 0, 1, 2, ...$, as shown in Fig.2. And let $S = \{s_1, s_2, s_3, ..., s_n\}$ be the finite set of possible states of the environment and $A = \{a_1, a_2, a_3, ..., a_n\}$ be the finite set of admissible actions the agent can take. At each time step t, the agent senses the current state $s_t = s \in S$ of its environment, and on that basis selects an action $a_t = a \in A$. As a result of its action, the agent receives an immediate reward r_t, and the environment's state changes to the new state $s_{t+1} = s' \in S$ with a transition probability $P_{ss'}(a)$.

The objective of the agent is to find an optimal policy $\pi^*(s) \in A$ for each state s to maximize the total amount of reward it receives over the long run. Q-Learning algorithm provides an efficient on-line approach to determine

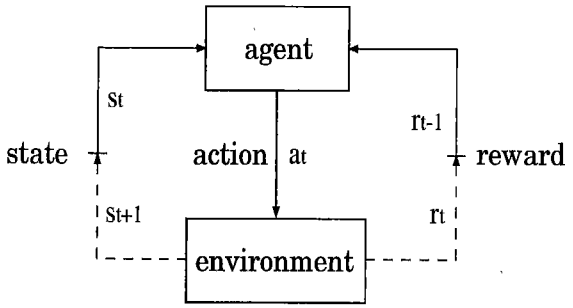


Fig. 2. An illustration of agent's interaction with its environment.

the optimal policy by estimating the optimal Q-values $Q^*(s, a)$ for pairs of states and admissible actions.

The Bellman optimality equation for $Q^*(s, a)$ is given as follows:

$$Q^*(s,a) = \sum_{s'} P_{ss'}(a)[R^a_{ss'} + \gamma \max_{a'} Q^*(s', a')] \quad \cdots\cdots (1)$$

where $R^a_{ss'} = r_t$ is the immediate reward from taking action a in the state s and transitioning from state s to s', a' is the admissible action in the new state s', and γ $(0 \leq \gamma \leq 1)$ is a scaling factor used to discount the future rewards. If γ is small, it means that the expected future rewards count for less.

Any policy selecting actions that are greedy with respect to the optimal Q-values is an optimal policy [19]. Thus, the optimal policy is

$$\pi^*(s) = arg \max_a (Q^*(s,a)) \quad \cdots\cdots (2)$$

Without knowing the $P_{ss'}(a)$, the Q-Learning algorithm can find the $Q^*(s, a)$ in a recursive manner by using the available information s_t, a_t, s_{t+1} and r_t. The update rule for Q-Learning is

$$Q_{t+1}(s,a) = \begin{cases} Q_t(s,a) + \alpha \Delta Q_t(s,a) \\ \qquad\qquad if\ s = s_t\ and\ a = a_t \\ Q_t(s,a) \quad otherwise \end{cases} \quad \cdots\cdots (3)$$

where α is the learning rate, $Q_t(s, a)$ and $Q_{t+1}(s, a)$ are the Q-value for state-action pair (s,a) at time t and $t+1$, respectively, and

$$\Delta Q_t(s,a) = \{r_t + \gamma \max_{a'}[Q_t(s_{t+1}, a')]\} - Q_t(s,a) \quad \cdots\cdots (4)$$

The learning rate α $(0 < \alpha \leq 1)$ reflects the degree to which estimated Q-values are updated by new data. High values imply more rapid updates, with a risk of instability [11].

If the Q-value for each admissible state-action pair (s,a) is visited infinitely often, and the learning rate α decreases over the time step t in a suitable way, then as $t \to \infty$, $Q_t(s, a)$ converges with probability one to $Q^*(s, a)$ for all admissible pairs (s, a).

3.2 **QL-based Supplier Bidding Strategy** In the repeated day-ahead electricity auction market, each supplier will attempt to maximize his/her profit in a long run and to reduce risks. The need to maximize profit and manage risks at the same time is becoming a dominant industry problem [13]. Based on the Q-Learning algorithm, a bidding strategy for suppliers is developed to balance the tradeoff between the expected profit and risks. To simplify the analysis, it is assumed that each supplier bids his/her maximum generation capacity as the bidding quantity at each hour of every trading day. Therefore, the bidding strategy results in a bidding price decision-making problem.

As described earlier, it is assumed that each supplier has information only on his/her own cost and the public information of hourly market price, but lacks of information on the rivals. Thus, the bidding process is a stochastic process. During this stochastic bidding process, each supplier will attempt to meet his/her objectives of:

- increasing his/her profit from day to day,
- satisfying the target utilization rate on his/her generator everyday,

as described in Ref. (12). The target utilization rate is defined as the ratio between the expected dispatched power (MWh) and the maximum power output (MWh) of generator everyday.

To apply the Q-Learning algorithm in developing the bidding strategy, it is necessary to define the states, actions, and rewards first.

1) *States*: The state of environment is represented by the market price, and has 20 different levels which is equally distributed between 0 \$/MWh and the market ceiling price that is specified to 20 \$/MWh here. The market ceiling price is a capped price which is designed to prevent unacceptable market outcomes, as did in the California's electricity market.

As shown in Fig.3, if the market price is within the interval of (19,20] \$/MWh, then the environment's state is in state 19. It should be noted that the state of environment should include other market information such as predict power load of the next trading day, but these are not taken into account here for simplicity.

2) *Actions*: Each rational supplier will generate bidding price between his/her unit production cost and the market ceiling price. Therefore, it is assumed here that each supplier's admissible actions are represented by 20 intervals which are equally distributed between his/her unit production cost and the market ceiling price, as shown in Fig.3. Applying an action a is to randomly generate a bidding price in the ath interval.

3) *Rewards*: Taking into account the requirement of utilization rate on supplier's generator, it is assumed here that it is required a constant target utilization rate on generator for each hour of everyday, for simplicity. Considering the target utilization rate for that hour, the reward $r_{i,h}(s,a)$ of supplier i from his bids at hour h under action a and state s is calculated based on the following formula, which can be viewed as a penalty function:

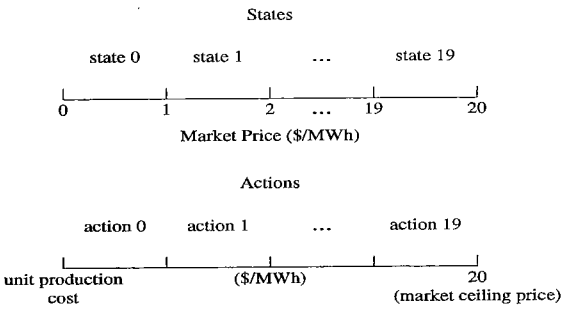


Fig. 3. The definition of states and agent's actions.

Table 1. An example of the Q-values of state - action pairs.

	...	action 11	action 12	action 13	...	action 16	...
...							
state 15		**512**	299	313		390	
state 16		395	**543**	489		294	
state 17		543	563	**608**		231	
state 18		489	561	600		**718**	
...							

$$r_{i,h}(s,a) = \pi(i,h) * (\frac{utl_a(h)}{utl_t})^n \quad \cdots\cdots\cdots\cdots (5)$$

where $\pi(i,h)$ is the payment of supplier i received for winning at hour h minus the costs of production, utl_t is the target utilization rate, $utl_a(h)$ is the actual utilization rate at hour h, and n ($n = 0, 1, 2, ...$) is a constant which shows how strictly a supplier tries to satisfy the requirement of utilization rate. If n is large, it implies that the supplier is strict in satisfying the requirement of utilization rate and can be thought to be of risk averse type. If $n = 0$, there is no penalty effect on the reward, and the supplier can be viewed to be opportunistic.

3.3 **Algorithm Implementation** We assume that whether the operation status of generator is on or off at any hour, it does not affect the operation status of generator at the next hour. Therefore, the whole day profit-maximizing problem can be decomposed into an hourly profit-maximizing problem. Based on this consideration, in this paper, Q-values for state-action pairs at each hour of a supplier are stored in a lookup table. An example of Q-values for state-action pairs is given in Table 1, where the Q-values in bold style are the maximum values under each state and the actions associated with them are the optimal actions a supplier would take most likely.

The steps of suppliers' learning and bidding are given as follows:

1) *Step 1: State identification.* At the beginning of the current trading day, each supplier uses the publicly available 24 separate hourly market prices on the previous trading day as the 24 hourly states of the current trading day.

2) *Step 2: Action selection.* After having obtained the 24 hourly states, each supplier inquires his/her Q-value lookup tables to select the optimal action with maximum Q-value in each state, and generates the bidding price

Table 2. The maximum capacities (MC), unit production costs (UPC) and startup cost of 10 agents in two supply cases.

	agent type	UPC ($/MWh)	MC (MW)	startup cost ($)	number of agents
supply case one	I	8.0	50	40	4
	II	10.0	50	40	3
	III	12.0	60	40	3
supply case two	I	8.0	50	40	6
	II	10.0	50	40	2
	III	12.0	60	40	2

at each hour according to the definition of an action.

To balance the exploration and exploitation of suppliers' learning from the dynamic electricity auction market, ϵ-greedy method [18] is introduced to the QL-based supplier bidding strategy. That is, during the action selection process, the supplier selects most of the time an action a with maximum $Q(s,a)$ in the state s; but, with a small probability ϵ, he also randomly selects an action a from all the admissible actions in the state s, independently of the Q-values $Q(s,a)$, to explore the new optimal bidding strategy in the dynamically competitive market.

3) *Step 3: Q-value update.* At the end of the current trading day, after being notified of the dispatched power and the market price at each hour, each supplier calculates the rewards according to (5), and updates the Q-values of each hour based on the available rewards and next states which are the hourly market prices of current trading day, according to (3), (4).

4. Simulation Results

An agent-based simulation method is developed here to test the bidding strategy proposed in the above section. The application of agent-based simulation method to deal with issues in the deregulated electricity industry is a newly promising research area [14] [15]. In this paper, it is assumed that there are two supply cases and in each case there are 10 adaptive agents, each of them representing a supplier who participates in the day-ahead electricity auction market and has the ability to explore and exploit the optimal bidding strategy to meet his/her profit-maximizing goal in the competitive environment. Table 2 gives the maximum generation capacities, unit production costs and startup costs of these agents. As can be seen from the capacity levels of all agents, no supplier possesses the market power since no agent has the dominant market share.

Many simulation cases have been carried out. In each case, to simulate that all initial Q-values of each agent at each hour are obtained from a long period of trading experience, the simulation process is designed to run first for 10,000 trading days to obtain these initial Q-values of agents. The discount factor γ of all agents is set to 0.1. The learning rate α is designed to be state-action dependent varying with time, as used in Ref. (16). That is, the learning rate $\alpha_{d,h}(s,a)$ of each agent at hour h on the trading day d is inversely proportional to the visited number $\beta_{d,h}(s,a)$ of state-action pair (s,a) up to

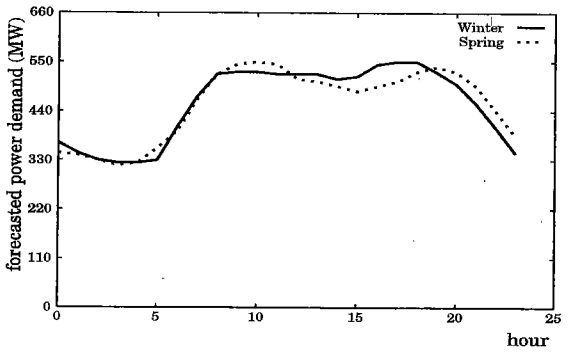


Fig. 4. The forecasted power demand in winter and spring by the ICA.

the present trading day, as follows:

$$\alpha_{d,h}(s,a) = \frac{1}{\beta_{d,h}(s,a)} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

After this learning process, the learned initial Q-values are used to develop optimal bidding strategy proposed in the Section 3.

4.1 Simulation Case 1 To firstly show the feasibility of the proposed bidding strategy, its impacts on the market price are investigated under discriminatory pricing rule. With the learned Q-values under winter power demand situation, the simulation process are carried out for 2,000 trading days with the strategic bidding of agents in supply case one. The forecasted power demand during the first 1,000 trading days is the winter power demand, and is changed to the spring power demand during the second 1,000 trading days, as shown in Fig.4. The parameters $\gamma = 0.1$, $\alpha = 0.5$, *target utilization rate* $= 0.75$, $\epsilon = 0.1$ and $n = 1$ are the same for all agents.

It should be pointed out that different parameters define different individual characters of agents – risk averse or opportunistic, non-greedy or greedy. But the impacts of individual character of each agent on his/her rewards and actual generator utilization rate are not investigated in this paper for concise reason.

Fig.5 shows the average hourly market price during the winter and spring. As shown in this figure, the intense competition among agents leads to the lowest market prices during the off-peak load periods such as at hour 2, 3 and so on, where the electricity supply is much bigger than the power demand. However, during the peak load periods such as at hour 16, 17 and 18 in winter, at which there are shortages of the power supply, the market prices are very close to the market ceiling price due to the agents' learning to fully take advantage of the market opportunity. These facts show that the proposed QL-based bidding strategy is successful in generating optimal bidding prices at different hours and in different seasons for agents in the day-ahead electricity auction market.

The agents' learning can be seen in Fig.6, which shows the market prices at hour 17 during the 2,000 trading days. Power load changes from 550 MW to 506 MW at hour 17 on the trading day 1,000, when the season

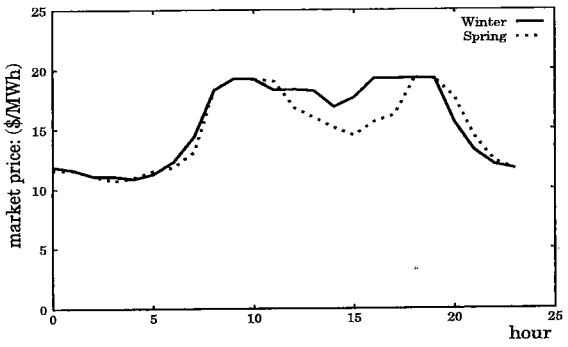


Fig. 5. The average hourly market price of everyday under discriminatory pricing rule.

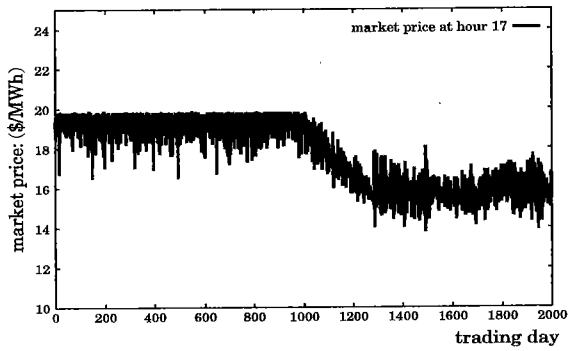


Fig. 6. The everyday market price at hour 17 during the 2,000 trading days.

changes from winter to spring. As can be seen from this figure, when the power load changes, the market price changes accordingly as adaptive agents interact with each other and learn from experience to develop their optimal bidding prices. It should be noted that, due to the slow convergence of Q-Learning algorithm, the market prices at hour 17 take a few days to reach a new dynamic equilibrium. But it is believed that the effect of this drawback can be reduced to some degree when each agent initializes his/her Q-values on a seasonal basis, takes actions and updates Q-values accordingly.

4.2 **Simulation Case 2** To further show its feasibility, the proposed bidding strategy is compared with another simple bidding strategy for suppliers, which is a simplified version of the "naive reinforcement learning algorithm" used in Ref. (12). The simple bidding strategy can be summarized as follows: if the supplier fails to achieve his/her target utilization rate on the generator at certain hour on the previous trading day, then subtracts a random percentage from the previous bidding price; otherwise, adds a random percentage to previous bidding price to create the next day's bidding price at that hour. The random percentage is generated from a uniform distribution with a range $\pm 10\%$ and a mean of 0.

In this simulation case, among the 10 agents in the supply case I, there are two of agent type I, II and III using the proposed Q-Learning based bidding strategy, respectively. These agents have the same parameters as those described above. All the others of each agent type use the simple bidding strategy. The auction market adopts the discriminatory pricing rule and has the forecasted winter power load. Simulation results are given in Table 3. As shown in this table, the proposed bidding strategy led to better rewards and actual generator utilization rate for suppliers from the auction market than the simple bidding strategy did.

Table 3. The average daily rewards and actual generator utilization rate of each agent type in supply case one using different bidding strategy in the auction market.

agent type	proposed strategy		simple strategy	
	Rewards ($)	Rate	Rewards ($)	Rate
I	9679	0.97	9439	0.81
II	7574	0.97	7459	0.81
III	6458	0.79	6205	0.74

Table 4. Impact of startup cost on their average daily rewards, bid prices and actual generator utilization rates as agents of type III are considered.

startup cost ($)	rewards ($)	bid prices ($/MWh)	actual generator utilization rates
0	5420	16.7	0.68
40	5340	16.7	0.67
80	5267	16.7	0.67
120	5151	16.7	0.67
160	5067	16.7	0.67

4.3 **Simulation Case 3** The impacts of startup cost on average daily rewards, bid price and actual generator utilization rate are studied in this case.

10 agents in supply case one with the same parameters used in simulation case 1 compete against each other under the discriminatory pricing rule and winter power demand. Table 4 shows the impact of startup costs as agents of type III are considered. According to this table, with an increase in startup cost, the average daily rewards decrease. However, the average daily bid prices and actual generator utilization rates of these agents remain unchanged (or little change). This can be explained that during the competition, the fixed Q-Learning parameters of each agent means the bidding strategy of each agent remains unchanged, therefore, with all bidding strategies fixed, each agent will bid prices in a certain range at each hour and the market will reach a dynamic equilibrium in a long term.

4.4 **Simulation Case 4** To investigate which market pricing rule is better for the auction market used in this paper, the average hourly market prices under two major pricing rules are compared. The 10 agents in supply case one are allowed to compete for 1,000 trading days with the forecasted winter power demand. Under the discriminatory pricing rule, all 10 agents bid strategically. However, under the uniform pricing rule, agents (type I) with the lowest costs bid their true costs everyday while others bid strategically. Q-Learning parameters of those bidding strategically are the same as those used above.

Fig.7 shows the comparison result. As can be seen from this figure, the market price under uniform pric-

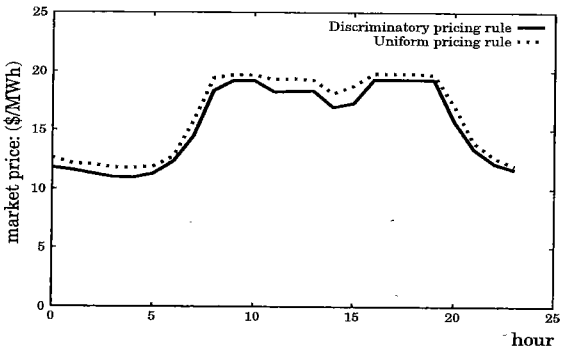


Fig. 7. The average hourly market prices in winter under two market pricing rules.

Table 5. The average daily rewards and actual generator utilization rate of each agent type in two supply cases when agents of type I bid strategically or bid their true costs.

	agent type	bid true cost		bid strategically	
		Rewards ($)	Rate	Rewards ($)	Rate
supply case one	I	9863	1.0	9726	0.97
	II	7199	0.94	7316	0.94
	III	5694	0.63	5768	0.67
supply case two	I	9769	1.0	9853	0.96
	II	7042	0.80	7348	0.86
	III	6067	0.60	6232	0.64

ing rule is higher than that under discriminatory pricing rule. This means that, in a competitive auction market where no supplier possesses the market power, uniform pricing rule could bring more profits to electricity suppliers, and that consumers could have to pay more under this market pricing rule. It also can be seen from this figure that, no matter which market pricing rule is adopted, market prices will decrease when competition degree is high and will rise close to the market ceiling price when competition degree is low.

4.5 **Simulation Case 5** In this case, whether or not it is better for agents (type I) with the lowest costs to bid their true costs under uniform pricing rule is investigated with the forecasted winter power demand. Supply case one and supply case two are used here. Agents of type II and III all bid strategically and have the same Q-Learning parameters as those used previously, so do agents of type I when they bid strategically.

The average daily rewards and actual generator utilization rate of each agent type are given in Table.5. From this table, it is better for agents of type I to bid their true costs under uniform pricing rule in supply case one, where any type of agents has not the dominant market share. Bidding their true costs in this case will eliminate risks and guarantee that their generators will be always called into operation. When they try to bid more than their true costs, they would face risks of their bids not being accepted and suffer profit decreases accordingly. At the same time, profit decreases by the agents of type I would lead to profit increases for other types of agents in the same competitive market, as can be seen from Table 5. However, simulation results also show that, in the supply case two where agents of type I have the dominant market share, bidding strategically could bring them more profits than bidding their true costs, although their actual generator utilization rates decrease to some degree. Moreover, all other agents in this supply case will benefit from the strategic bidding behavior of agents of type I.

5. Conclusions

Based on the Q-Learning algorithm, an optimal bidding strategy was proposed in this paper to provide suppliers an optimal approach to maximize their profits in the long run from the day-ahead electricity auction market. Each supplier with the proposed bidding strategy can learn from experience and make full use of the public information of the market. A penalty function was introduced to the calculation of the reward from supplier's bids. The impacts of the proposed bidding strategy on the market price were analyzed. The proposed QL-based bidding strategy was compared with a simple bidding strategy. Simulation results have shown the feasibility of this QL-based supplier bidding strategy.

In this paper, it has been shown that uniform pricing rule could lead to a higher market price than discriminatory rule. Although the purpose of this paper is not intended to deal with the market design, it still provided a deep insight into the complex new electricity markets.

Also, it has been shown that whether or not agents with the lowest costs should bid their true costs depends on the market supply condition. When agents with the lowest costs have a dominant market share, they could benefit from their strategic bidding behaviors.

We realize that the proposed bidding strategy is still some kind of simple at the current stage and a practical strategy should take many constraints and conditions into consideration, such as minimum output, ramp rate of generators. More work will be done in the future to make it a practical method.

(Manuscript received Aug. 23, 2002,
revised Dec. 12, 2002)

References

(1) Harry Singh: IEEE Tutorial on Game Theory Application in Electric Power Markets, IEEE Power Engineering Society, Winter Meeting, New York (1999)

(2) D.Fudenberg and J.Tirole: Game Theory, Cambridge, Massachusetts: The MIT Press (1991)

(3) R.W.Ferrero, J.F. Rivera, and S.M.Shahidehpour: "Application of games with incomplete information for pricing electricity in deregulated power pools", *IEEE Trans. on Power Syst.*, Vol.13, No.1, pp.184-189, Feb. (1998)

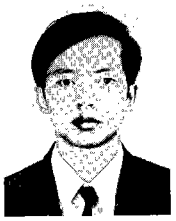
(4) F.S.Wen, and A.K.David: "Oligopoly Electricity Market Production under Incomplete Information", *IEEE Power Engineer Review*, pp.58-61, April (2001)

(5) K.Seeley, J.Lawarree, and C.C.Liu: "Analysis of Electricity Market Rules and Their Effects on Strategic Behavior in a Noncongestive Grid", *IEEE Trans. on Power Systems*, Vol.15, No.1, pp.157-162, Feb. (2000)

(6) C.W.Richter,Jr and G.B.Sheble: "Genetic Algorithm Evolution of Utility Bidding Strategies for the Competitive Marketplace", *IEEE Trans. on Power Systems*, Vol.13, No.1, pp.256-261, Feb. (1998)

(7) H.L.Song, C.C.Liu, J.Lawarree, and R.W.Dahlgren: "Optimal Electricity Supply Bidding by Markov Decision Process",

IEEE Trans. on Power Systems, Vol.15, No.2, pp.618-624, May (2000)
(8) F.S.Wen and A.K.David: "Optimal Bidding Strategies and Modeling of Imperfect Information Among Competitive Generators", *IEEE Trans. on Power Systems*, Vol.16, No.1, pp.15-21, Feb. (2001)
(9) Derek W.Bunn and Fernando S.Oliveria: "Agent-Based Simulation – An Application to the New Electricity Trading Arrangements of England and Wales", *IEEE Trans. on Evolutionary Computation*, Vol.5, No.5, Oct. (2001)
(10) S.Y.Hao: "A Study of Basic Bidding Strategy in Clearing Pricing Auctions", *IEEE Trans. on Power Systems*, Vol.15, No.3, pp.975-980, Aug. (2000)
(11) SEPIA's web site, http://www.htc.honeywell.com/projects/sepia
(12) J. Bower and D. Bunn: "Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the England and Wales electricity market", *J. Economic Dynamics and Control*, Vol.25, pp.561-592, March 2001.
(13) M.Ilic, and P.Skantze: "Electric Power Systems Operation by Decision and Control", *IEEE Control Syst. Magazine*, Vol.20, No.4, pp.25-39, Aug. (2000)
(14) C.C.Liu, J.H.Jung, G.T.Heydt V.Vittal, and A.G.Phadke: "The Strategic Power Infrastructure Defense (SPID) System", *IEEE Control Systems Magazine*, Vol.20, No.4, pp.40-52, Aug. (2000)
(15) S.A.Harp, A.Brignone, B.F.Wollenberg, and T.Samad: "SEPIA: A Simulator for Electric Power Industry Agents", *IEEE Control Syst. Magazine*, Vol.20, No.4, pp.53-69, Aug. (2000)
(16) Junhong Nie and Simon Haykin: "A Dynamic Channel Assignment Policy Through Q-Learning", *IEEE Trans. Neural Networks*, Vol. 10, No. 6, pp. 1443-1455, Nov. (1999)
(17) Arthur R.Bergen, Vijay Vittal: Power Systems Analysis, 2nd ed., Prentice-Hall (2000)
(18) Richard S. Sutton and Andrew G. Barto: Reinforcement Learning: An Introduction, Cambridge, MA:MIT Press (1998)
(19) A. G. Barto, S. J. Bradtke, and S. P. Singh: "Learning to act using real-time dynamic programming", *Artificial Intelligence*, Vol. 72, pp. 81-138 (1995)
(20) C. J. C. H. Watkins, Learning from Delayed Rewards, PhD thesis, Cambridge University, Cambridge, England (1989)



Gaofeng Xiong (Student Member) received his BS, MS degrees in Electrical Engineering from Hunan University, P.R.China, in 1992 and 1995, respectively. He was a faculty member at College of Mechanical and Automotive Engineering, Hunan University, from 1995 to 1999. He is currently a Ph.D. student at Electrical Engineering, Nagoya University, Japan. His research interests include power system economics, evolutionary computation, and reinforcement learning. He is a student member of the IEE of Japan, and the IEEE.

Tomonori Hashiyama (Member) received his Ph.D degree in Electrical Engineering from Nagoya University, Japan, in 1996. He is currently an associate professor at the Institute of Natural Sciences, Nagoya City University. His research interest include evolutionary computation, optimization, and intelligent system applications. He is a member of the IEE of Japan, and the IEEE.

Shigeru Okuma (Member) received his M.E. degree in systems engineering from Case Western Reserve University, OH, U.S.A, and his Ph.D. degree in Electrical Engineering from Nagoya University, Japan, in 1974 and 1978, respectively. Since 1990, he has been a Professor of Electrical Engineering at Nagoya University. His research interests are in the areas of power electronics, robotics, and evolutionary computation. He is a member of the IEE of Japan, and the IEEE.