Agent-Based Analysis of Capacity Withholding and Tacit Collusion in Electricity Markets

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

***Abstract*—This paper employs agent-based simulation to study energy market performance and, in particular, capacity withholding and the emergence of tacit collusion among the market participants. The energy market is formulated as a repeated game, where each stage game corresponds to an hourly energy auction. Each hourly energy auction is cleared using locational marginal pricing. Generators 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 and eight competing generator-agents, demonstrate the development of tacit collusion among generators even under competitive conditions.**

***Index Terms*—Agent-based simulation, capacity withholding, collusion, reinforcement learning, repeated games.**

NOMENCLATURE

$k, (\mathcal{K})$	Index (set) of buses.
$g, (\mathcal{G})$	Index (set) of generators.
$f, (\mathcal{F})$	Index (set) of consumers.
km	Line from bus k to bus m index.
$k(g)$	Bus on which generator g is connected.
$P_g^{\max}$	Generator g net capacity.
mc_g	Generator g marginal cost.
P_g	Generator g offer quantity in the spot market.
b_g	Generator g offer price in the spot market.
p_g	Unit g scheduled quantity in the spot market.
$\mathbf{p}$	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 $\mathrm{K} \cdot \mathrm{F}$).

Manuscript received September 19, 2006; revised April 3, 2007. This work was supported by the National Fellowship Foundation of Greece. Paper no. TPWRS-00646-2006.

The authors are with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece (e-mail: atellidou@eng.auth.gr; bakiana@eng.auth.gr).

Digital Object Identifier 10.1109/TPWRS.2007.907533

I. INTRODUCTION

THE liberalization of the energy sector allows for the operation of open markets where participants can choose among different products and different time horizons in order to conduct their trading agreements. In this new regime, electricity trade is usually conducted in three different kinds of markets, regarding the time horizon. In the long-term, where participants have the opportunity to come to direct agreement, bilateral contracts dominate. In the short term, two kinds of markets have prevailed: the day-ahead market, which is a forward market for next day delivery in one-hour or half-hour trading intervals, and the real-time or spot market, which serves as balancing mechanism for the next hour or half-hour. Both short-term markets are usually organized as auctions repeated every day (day-ahead market) or every several hours (real-time market).

In [1], Rothkopf deals with the important issue of daily repetition in electricity auctions. He states that the experience in auctions suggests that the repetition of auctions involving the same parties can have major consequences and argues that models of how auctions work in isolation may not predict well how they work when repeated daily. Both the single-stage and the iterated prisoner's dilemma are presented in order to illustrate how repetition of a game can lead to cooperation, when single-play model of the game predicts fierce competition. In [2], Axelrod investigates the conditions under which cooperation may emerge in a world of egoists without central authority. He states that the two-person prisoner's dilemma captures an important part of the strategic interaction and that what makes the emergence of cooperation possible is the possibility that the interaction will continue. With an infinite number of interactions, where the players can communicate only through the sequence of their behavior, one can expect the emergence of cooperation under certain conditions.

In [3], according to [1], Smith looks at auctions from the point of view of sociology. He gives great insight into the behavior of participants in regularly repeated auctions. Smith is convincing in arguing that bidders in a repetitive auction process form a special group with its own norms and behavior designed to protect the group's interests. Bidders in electricity auctions interact daily. While they do so at a distance, they are likely to form a social group that protects its own interests. This implies that the behavior studied by Smith will occur and should be taken into account in the design of electricity auctions.

Hence, one of the most important effects of the daily repetition of electricity auctions is the emergence of collusion. Collusion refers to combinations, conspiracies or agreements among sellers to raise or fix prices and to reduce output in order to increase profits [4]. Collusion does not necessarily have to involve an explicit agreement or communication between firms;

they may take their rivals' actions into account and coordinate their actions as if they were a cartel without an explicit or overt agreement. Such coordinated behavior is often referred to as *tacit collusion* or *conscious parallelism* [4]. Tacit collusion—in contrast to the explicit one—can be unstable since the participants can get higher short-term payoff by deviating from the collusive strategy; hence a myopic decision-making participant tends to deviate from the collusive behavior [5].

Although, as noted in [6], many of the attributes that facilitate collusion are present in electricity markets, little work has been done so far to study this. In [7], an infinitely repeated game of capacity-constrained price competition among symmetric firms is analyzed; in every period, the firms submit offers that specify the minimum price at which they are willing to supply their output up to capacity. The analysis focuses on the comparison of the level of collusion under the two prevailing pricing rules in electricity auctions, uniform and discriminatory pricing. In [8], the above results are extended to the case where firms may submit bids that are step functions of price-quantity pairs with any finite number of price steps. The analysis of the feasibility of collusion is limited to the uniform pricing auction, but two different rules are used to determine the uniform price. In [9], Oren deals with the problem of implicit collusion in congested electricity system from a game-theoretic point of view. His objective is to illustrate the inefficiency of passive transmission rights, since they may be preempted by the strategic bidding of generators. Hence, using a Cournot model of competition in a congested transmission network, he concludes that, "although none of the generators have market power by any measure of concentration, rational expectation of congestion leads to implicit collusion." In [5], a co-evolutionary genetic algorithm is used to model the strategies of the players, who participate in an electricity market formulated as a dynamic game with trading interactions being repeated over time. The results indicate that in a simple market two agents can "learn" to develop collusion in order to increase the market price.

An alternative to the game theoretic approach, used to study the behavior of market participants, is agent-based simulation (ABS) [10]. In the implemented ABS, the market participants develop their profit-maximizing strategy through the repetition of the game and reinforcement learning. Q-Learning [11] is among the most popular reinforcement learning algorithms, owing to its simplicity. Its main advantages are that it can be used online and it is model free—it does not need an explicit model of its environment.

Q-Learning was initially developed concerning the interactions of one agent in a static environment; in this form it has been used for the modeling of the participants' behavior in electricity markets [12]. However, as the interest for multi-agent interactions has increased, many extensions of the Q-learning algorithm have been proposed [13], [14]. Two important technical reports [15] and [16] enlighten the research on multi-agent reinforcement learning by presenting the relative literature and summarizing algorithms and convergence proofs.

Based on the extensions of Q-learning algorithm discussed above, Krause *et al.* analyze the strategic bidding in power markets in [17] and [18]. They present an extended agent-based study of an electricity auction, simulated as a repeated game. Their cases include both active players and agents who simply offer their marginal cost, underlining the difference in their behavior and their profits. The offer curves are simulated with two different ways, as a single pair of quantity-price and as linear function. Their analysis concludes that agents following a Q-learning strategy can learn to behave as game theory suggests.

In this paper, the power market operation is formulated as a repeated game with each stage being an hourly auction; Q-learning algorithm, in the form presented in [17] and [18], is used to model the bidding strategy of generators in this auction. The main difference is the followed policy; instead of the greedy one, in this paper a different approach based on the Metropolis criterion of simulated annealing is used. Our purpose is to study the behavior of the players in a network constrained market and the development of tacit collusion through capacity withholding. Section II presents the spot electricity market structure. Section III introduces the repeated games and discusses their analogies to electricity market. Section IV presents the fundamentals on reinforcement learning along with a description of the SA-Q-learning algorithm and a model of generator's behavior. Section V reports the results of our tests. Finally, Section VI summarizes our conclusions.

II. Electricity Market Structure

This paper studies the behavior of generators in a spot market with hourly trading interval. Generators submit energy offers to the independent system operator (ISO), declaring the power they are willing to sell at or above 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, $0 \le P_g \le P_g^{\max}$, 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, $\mathrm{mc}_g \le b_g \le pc$.

A. ISO Market Clearing Problem

The ISO collects the energy offers (P_g, b_g) of all generators, $g \in \mathcal{G}$, and based on the most recent forecast of the nodal demands, $\mathbf{d}$, computes the quantities, $p_g, \forall g \in \mathcal{G}$, and nodal prices, $\mathrm{LMP}_k, \forall k \in \mathcal{K}$, that clear the market, by solving the following optimal power flow (OPF) problem

$$\mathrm{Min} \sum_{g \in G} b_g \cdot p_g \tag{1}$$

subject to

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

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

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

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

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

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

limits, and (5) represent the unit active power output limits. The Lagrange multipliers of the nodal active power balance constraints (2) are the nodal prices (LMPs).

B. 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)

$$\text{Max Profit}_g = \left(\text{LMP}_{k(g)} - \text{mc}_g\right) \cdot p_g \tag{6}$$

subject to the following constraints:

$$0 \le P_g \le P_g^{\max} \tag{7}$$

$$\text{mc}_g \le b_g \le pc \tag{8}$$

as well as the ISO market clearing problem solution (1)–(5), needed to define p_g and $\text{LMP}_{k(g)}$ used in (6).

However, the generator does not have the information on the transmission network, the consumer demand 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 (LMP) information.

III. Repeated Games

A. Static Games

The essential elements of a game are:

1) the participants, who comprise the *set of players*;
2) the set of alternative choices each participant has, which is called the *set of actions*;
3) the *payoff* each participant gets for each outcome of the game (or combination of participants' actions) [19].

Let us assume the following n -player simple game:

1) Players 1 through n simultaneously choose actions $a_1 \in \mathcal{A}_1$ through $a_n \in \mathcal{A}_n$, respectively.
2) They receive their payoffs $u_1(a_1, \ldots, a_n)$ through $u_n(a_1, \ldots, a_n)$.

$\Gamma = \{\mathcal{A}_1, \ldots, \mathcal{A}_n; u_1, \ldots, u_n\}$ denotes the above static game and will be called stage game of the corresponding repeated game [20].

According to the above, the market operation, as described in Section II, can be seen as a stage game, where $\mathcal{G}$ is the set of players; then each generator can be regarded as player and his action space consists of all the possible selections of his energy offer, (P_g, b_g), i.e., his action space is the Cartesian product of the regions defined by constraints (7) and (8).

B. Infinitely Repeated Games

1) 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 tth 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 [20].

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

$$u^{(1)} + \delta u^{(2)} + \delta^2 u^{(3)} + \cdots = \sum_{t=1}^{\infty} \delta^{t-1} u^{(t)}. \tag{9}$$

The discount factor reflects the time value of money [20]; 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, as it will be justified in the discussion of the results.

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. Generator's Behavior Under Modified Q-Learning

A. Reinforcement Learning

Many learning theories have been developed as a result of man's effort to analyze the behavior of animals and artificial systems. Reinforcement learning (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; it studies complex behaviors, where sometimes taking an unpleasant action may lead to a long-term reward [21].

RL theory's basic elements are:

- the learner or the decision-maker, called the *agent*;
- everything it interacts with, called the *environment*.

The effects of agent's actions cannot be fully predicted; thus he must monitor his environment frequently and react appropriately, in order to learn from the consequences of his actions [22]. The basic concept behind RL is *trial-and-error* search, since the agent explores his environment and learns from his mistakes.

Recently, the research focused to multi-agent RL. Instead of a static environment, unknown but predictable to some extent—after some learning procedure—the agent has to face a constantly evolving environment containing other agents. In particular, the behavior of the other agents may change, as they also learn to better perform their tasks. This type of multi-agent non-stationary world creates a difficult problem for learning to act in these environments [16]. From a technical point of view, the research has been redirected from the realm of Markov decision problems to the one of game theory [15].

It is important to note that in such a multi-agent environment each agent does not interact with the others explicitly; its actions are not directed to the rest of the agents, but to his environment, so he does not try to teach the others. In RL, experience is the only teacher [23]; hence, an agent influences the behavior of

his competitors only implicitly, through the change he causes in their environment.

B. Q-Learning Algorithm

The Q-Learning algorithm, proposed by Watkins [11], is one of the most commonly used RL algorithms, because of its simplicity. Its main advantages are that it can be used online 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; thus it is a very useful tool for solving Markov decision problems. In order for the algorithm to be suitable for our game-theoretic multi-agent approach, some modifications—as presented in [17] and [18]—of the original algorithm, concerning the Q values, have been adopted.

The agent g in Q-Learning keeps in memory a function $Q_g(a_g)$[2] that represents the expected payoff he believes he will obtain by taking an action a_g; the function of the expected payoff is represented by a one-dimensional lookup table indexed by actions, whose elements are defined as Q-values.

The agent's experience, concerning his interaction with the environment, consists of a sequence of distinct stages. Let $\mathcal{A}_g = \{a_{g,1}, a_{g,2}, \ldots, a_{g,A_g}\}$ be the set of A_g possible actions the agent g can take. In the tth stage, the agent:

1) selects and performs an action $a_g^{(t)} \in \mathcal{A}_g$ using a policy;
2) receives an immediate payoff $u_g^{(t)}(a_g)$;
3) updates his Q values according to (10) at the bottom of the page.

According to (10), only Q values corresponding to the last action chosen are updated. $\alpha_g^{(t)}(a_g)$ 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 [24], [25].

C. SA-Q-Learning Algorithm

The SA-Q-Learning algorithm was proposed by Guo *et al.* [26] 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 [27], used in the SA algorithm [28], in order to determine the action-selection 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 $a_{g,r} \in \mathcal{A}_g$ randomly.
b) Selects an action $a_{g,p} \in \mathcal{A}_g$ following a greedy[3] policy:

$$a_{g,p} = \arg\max_{a_g} Q_g^{(t-1)}(a_g).$$

c) Generates a random number $\xi \in (0, 1)$.
d) Selects and performs action $a_g^{(t)} \in \mathcal{A}_g$ as follows:

$$a_g^{(t)} = \begin{cases} a_{g,p}, & \text{if } \xi \geq \exp\left[\frac{Q_g^{(t-1)}(a_{g,r}) - Q_g^{(t-1)}(a_{g,p})}{T_g^{(t)}}\right] \\ a_{g,r}, & \text{otherwise.} \end{cases} \tag{11}$$

e) Calculates $T_g^{(t+1)}$ 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 [26]. Let $T_g^{(t)}$ be the temperature in the tth 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

$$T_g^{(t+1)} = \lambda T_g^{(t)}, t = 0, 1, 2, \ldots. \tag{12}$$

D. Modeling Generator's Behavior

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

1) Action: The generator-agent action is the selection of the offer quantity and price, (P_g, b_g). The agent's action space is discretized, by discretizing both the offer quantity and the offer price intervals of variation (7), (8) into $\mathcal{A}_g^P$ and $\mathcal{A}_g^b$ levels, respectively.

2) Payoff: The payoff received by each agent during an auction round is equal to the profit, in €, the agent makes by participating in the spot market, defined in (6).

The learning rate is designed to be action dependent, as in [24] and [25]. The learning rate $\alpha_g^{(t)}(a_g)$ is inversely proportional to

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

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

$$Q_g^{(t)}(a_g) = \begin{cases} \left(1 - \alpha_g^{(t)}(a_g)\right) Q_g^{(t-1)}(a_g) + \alpha_g^{(t)}(a_g) u_g^{(t)}(a_g), & \text{if } a_g = a_g^{(t)}, \\ Q_g^{(t-1)}(a_g), & \text{otherwise} \end{cases} \tag{10}$$

the visited number $\beta_g^{(t)}(a_g)$ of action a_g up to the present trading stage, as follows:

$$\alpha_g^{(t)}(a_g) = \frac{1}{\beta_g^{(t)}(a_g)}. \quad (13)$$

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 generator-agents select an action $a_g^{(t)} \equiv (P_g^{(t)}, b_g^{(t)})$ according to the policy defined by the SA-Q Learning (11), and submit their energy offers defined by the selected action to the ISO.

Step 2) The ISO processes the energy offers submitted by all generator-agents, along with transmission system and nodal demand information, and computes the quantities and prices that clear the market by solving (1)–(5).

Step 3) The ISO posts the public information on nodal prices, $\mathrm{LMP}_k, \forall k \in \mathcal{K}$, and informs every generator-agent $g \in \mathcal{G}$ about the quantity, p_g, of his energy offer accepted in the spot market.

Step 4) All generator-agents use the information they receive from the ISO (Step 3) to compute profits and update their Q Tables according to (10).

Step 5) The stage count, t, is updated; the temperature is updated according to (12); the learning rate is updated according to (13). The whole process, Step 1 through Step 5, is repeated if the stage count, t, is less than the maximum number of stages, $t^{\max}$.

In the ISO market clearing process (step 2), if the energy prices of different energy offers arithmetically coincide and the respective quantities of such offers are not included in their entirety in the ISO schedule, then the specific energy offers to be partially or wholly included in the ISO schedule are selected at random. In this way, asymmetries in the simulation results caused by a consistent selection of the winning energy offers among the marginal energy offers with identical offer prices are avoided.

F. Model Limitations

The social interactions and the human behavior are rather complicated to be modeled through the simple algorithm described before. Hence, it should be noted that the objective of the simulation is not the realistic representation of the market environment, but the analysis of some undesirable—for the effective operation of the market—phenomena that can be observed even in this simple model of the market procedures and the participants' behavior. The basic assumptions of our model are the following.

The price elasticity of demand is zero. It is well recognized [29] that "electricity balancing markets have little or no demand elasticity"; hence the hourly market that we simulate is modeled more realistically when there is no demand elasticity. This is an assumption also made in [30], where the market power in a congested network is studied and the model used considers inelastic system load. In order to avoid arbitrarily high prices, all generators' offer prices are subject to a price cap.

Fig. 1. Two-node test system.

TABLE I
GENERATION DATA FOR CASES A AND B

Generators g	$P_g^{\max}$ [MW]	mc_g [€/MWh]
Gen-1	500	20
Gen-2	500	30

Furthermore, note that in all stage game repetitions the nodal load demand vector and the transmission system conditions (line status, parameters, and limits) remain the same. Although our assumption may not be realistic, our objective is not to simulate the real-world conditions but to study if the agents are able to learn to respond to a specific environment.

Finally, only the LMPs are considered as public information, while the marginal costs, the bids and the dispatched quantities of the rivals as well as the transmission network are not publicly known. Hence, each generator-agent does not know anything about the other participants and there is no explicit interaction with them. The rivals as well as the transmission network are embedded in the environment of each generator-agent and are treated as part of it.

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. Two 500-MW generators, who represent the market players, compete to serve the two constant loads shown in Fig. 1. The generator data (net capacity, $P_g^{\max}$, and marginal cost, mc_g) are shown in Table I. 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.

In Case B, each generator participates in a repeated energy auction trying to maximize its profits by reinforcement learning, as described in Section V.

In Case C, each of the two generators of Table I is "split" into four identical competing generators as shown in Table II. By increasing the number of competitors from two to eight in this case the level of competition is increased (compared to Case B) while the remaining data (total generation and transmission capacity and total demand) are the same.

Parameter Selection. All generators 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

TABLE II
GENERATION DATA FOR CASE C

Generators g	P_g^{max} [MW]	mc_g [€/MWh]
Gen-1a, Gen-1b, Gen-1c, Gen-1d	125	20
Gen-2a, Gen-2b, Gen-2c, Gen-2d	125	30

TABLE III
MARKET CLEARING UNDER COMPETITIVE PRICES (CASE A)

Nodes k	Generators g	P_g^{disp} [MW]	LMP_k [€/MWh]	$Profit_g$ [€]
Node-1	Gen-1	200	20	0
Node-2	Gen-2	100	30	0

temperature $T^{(0)}$ and the constant λ of the temperature-dropping criterion. All generators have the same parameters $T^{(0)} = 100\,000$ and $\lambda = 0.99$.

In the simulations presented, each agent's action space is discretized. The step for the offer price has been set equal to 2€/MWh, while the step for the offer quantity has been set equal to 10 MW for Case B and 2.5 MW for Case C.

A. Reference Case: Two Generator-Agents

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 III. Owing to the 100-MW transmission limit, the cheaper generator, Gen-1, is dispatched only up to 200 MW (100 MW serve the local Node-1 demand, while the remaining 100 MW are transported to Node-2, congesting the transmission line). 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, since no generation capacity limit is active in the OPF solution, the LMP at each node is equal to the marginal cost of the local generator. Hence, neither generator makes profit from the energy market, while the ISO collects €1000 of congestion rent.

B. Oligopoly Case: Two Generator-Agents

If both generator-agents act strategically, trying to maximize profits through reinforcement learning, the resulting market conditions are shown in Figs. 2 and 3, where the results of the last 500 stages out of the total 2500 stages of the SA-Q learning algorithm are presented. As shown in Fig. 2 the first agent withholds capacity by offering only 200 MW, in order to leave the transmission line uncongested and be paid at the LMP of Node-2, thus increasing profits. According to [9], this is consistent with Coase Theorem (1960), which supports the argument that "in the absence of transaction costs and with public knowledge of transmission capacity, bargaining among buyers and sellers will capture all congestion rents." In our case, despite the fact that Gen-1 does not know the transmission capacity, he "learns" to withhold capacity through repetition.

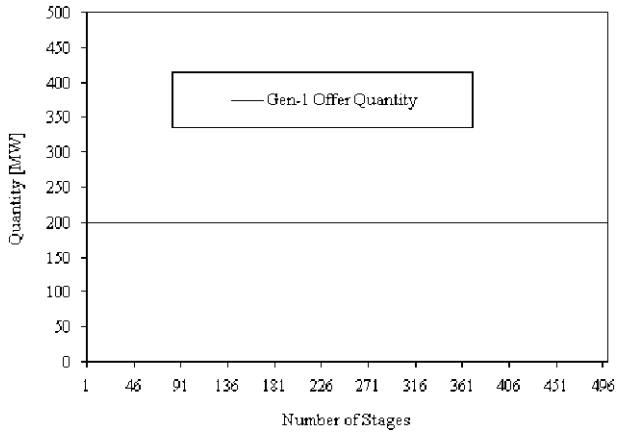

Fig. 2. Generator 1 offer quantity, equal to Generator 1 dispatched quantity.

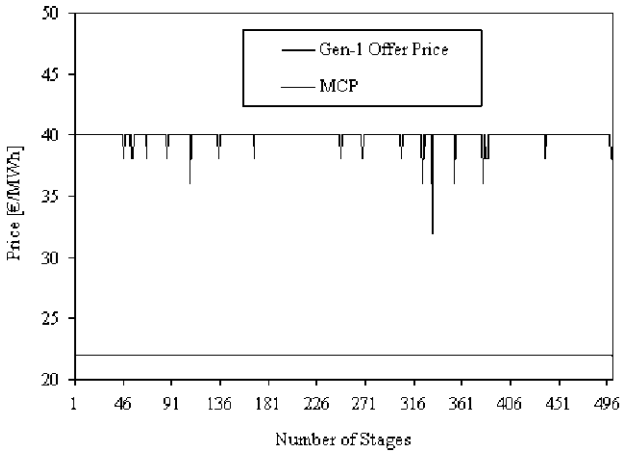

Fig. 3. Gen-1 offer price and market clearing price.

Since there is no congestion, both producers are paid at the same system-wide market clearing price (MCP). The resulting MCP is very close to the market price cap, as shown in Fig. 3 owing to the fact that agent Gen-2 realizes his monopoly power over the last 100 MW of the local, Node-2, demand and the absence of demand elasticity at Node-2.

The above outcome is a Nash equilibrium since no agent can profitably deviate as can be easily shown.

C. Increased Competition Case: Eight Generator-Agents

Here, each of the two generators of Case B is "split" into four identical competing generators in order to increase competition among generators. Fig. 4, where the results of the last 500 stages out of the total 5000 stages of the SA-Q Learning algorithm are presented, shows that the generator-agents of Node-1 collectively withhold capacity, even though there is no communication amongst them. The generators located on Node-1 seem to develop some kind of cooperation through the learning procedure, and keep their cumulative offer quantity below the threshold of 200 MW, managing to prevent the transmission line congestion.

From the game theoretic point of view, this developed cooperation is consistent with Friedman's (1971) Theorem [20] for infinitely repeated games, which suggests that the outcome of an infinitely repeated game may not be a Nash equilibrium of the

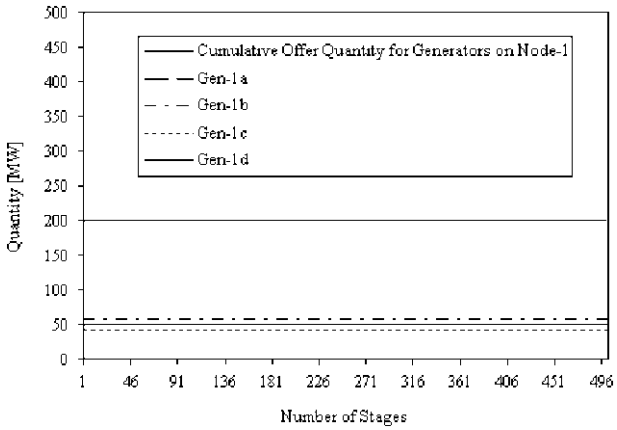

Fig. 4. Cumulative offer quantity of generators on Node-1, equal to their cumulative dispatched quantity.

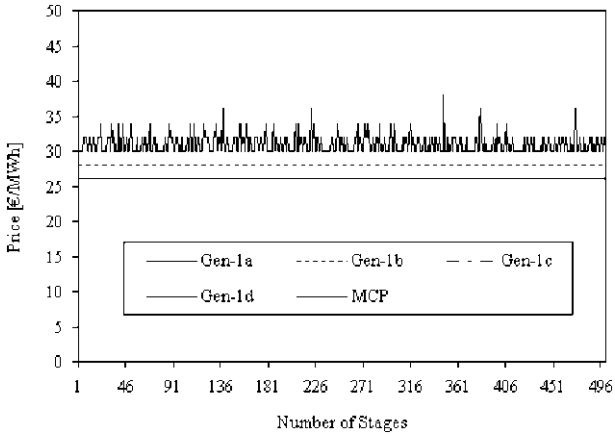

Fig. 5. Offer Prices of Generators on Node-1 and system-wide LMP.

corresponding stage game. In [20], the infinitely repeated Prisoner's Dilemma is analyzed; it is suggested that the outcome of this infinitely repeated game can be the choice of cooperation for both players, which is not a Nash equilibrium of the stage game, and yet it is a subgame-perfect Nash equilibrium. In case one generator deviates from the collusive strategy, he may end up gaining increased short-term profit (just for the stage he deviates); this would cause more aggressive bidding from the other generators and his profit would be lower in the new competitive environment.

At this point, it is important to discuss the development of cooperation, in the absence of the discount factor in our model. In [31], Axelrod states that "an important conclusion drawn from this investigation is that foresight is not necessary for the evolution of cooperation." Reference [31] analyses the conditions in a variety of social incidences, under which the emergence of cooperation is possible. Commenting the cases of trial-and-error learning, like RL, Axelrod concludes that "the players can come to cooperate with each other through trial-and-error learning about possibilities for mutual rewards . . . even through a blind process of selection of the more successful strategies with a weeding out of the less successful ones." But, since learning through trial-and-error is "slow and painful," foresight is something that can be used in order to "speed-up the evolution of cooperation." Hence, discount factor is not necessary for the emergence of cooperation, but it can play an accelerating role.

Since there is no LMP difference, all generators are paid at the system-wide MCP. In contrast to the previous case, as can be seen in Fig. 5, the MCP is not very high; in fact, it is close to the marginal cost of the generators of Node-2. This can be easily explained by the increased competition among four competing generators in Node-2. The generator-agents of Node-1 offer prices below 30 €/MWh, ensuring the dispatch of their offered quantity (collectively 200 MW); the four generator-agents of Node-2 have to compete for supplying the remaining 100 MW of Node-2 demand. The outcome of the developed competition is the low level of the MCP. The observed fluctuation in MCP (Fig. 5) can be attributed to the fact that the generator-agents of Node-2 try to raise their profit either by increasing their dispatched quantity, thus lowering their bid price, or by increasing the MCP, thus bidding a higher price at the expense of dispatched quantity.

TABLE IV
EXECUTION TIMES

Case	No of Stages	Execution Time (min)
Case B	2,500	21.65
Case C	5,000	47.84

It is important to discuss why identical agents with the same parameters do not develop the same optimal policy. The reason is that they face different environments. If each one of these agents had been placed separately in the same environment, then they would have generated the same offers. In our case, the first iterations of the algorithm are purely exploratory and during exploration the actions are selected randomly and are different for each player. Hence, each agent ends up facing his own environment—slightly different from the ones of his co-players—despite the symmetry among them. Consequently, everyone develops his own policy and based on that generates his offers.

D. Simulation Environment-Execution Times

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. Table IV presents execution times for test cases B and C.

VI. CONCLUSION

An analysis of the development of tacit collusion and capacity withholding 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 was modeled as an adaptive agent, following a SA-Q-learning bidding behavior. Test cases on a simple two-node test system, with two and eight competing generators led to the following conclusions.

Under high market concentration (Case B) generators participating in a repeated energy auction can learn to develop capacity withholding strategies (Gen-1) and to recognize their locational

market power (Gen-2) based only on publicly available (LMP) information.

Even under competitive conditions (Case C) generators participating in a repeated energy auction can learn to develop tacit collusion, in order to capture the ISO's congestion rents. It is important to note that tacit collusion arises, even though there is no teaching (i.e., punishing aggressive behavior by competitors) in the learning algorithm.

REFERENCES

[1] M. H. Rothkopf, "Daily repetition: A neglected factor in the analysis of electricity auctionsr," *Elect. J.*, vol. 12, no. 3, pp. 60–70, Apr. 1999.
[2] R. Axelrod, "The emergence of cooperation among egoists," *Amer. Pol. Sci. Rev.*, vol. 75, no. 2, pp. 306–318, Jun. 1981.
[3] C. W. Smith, *Auctions: The Social Construction of Value*. Berkeley, CA: Univ. California Press, 1990.
[4] R. S. Khemani and D. M. Shapiro, "Glossary of industrial organisation economics and competition law," 1993, commissioned by the Directorate for Financial, Fiscal and Enterprise Affairs, OECD. [Online]. Available: http://www.oecd.org/dataoecd/8/61/2376087.pdf.
[5] T. D. H. Cau and E. J. Anderson, "A co-evolutionary approach to the tacit collusion of generators in oligopolistic electricity markets: Price-wise linear bidding structure case," *Proc. 2003 Congr. Evolutionary Computation, CEC'03*, vol. 4, pp. 2306–2313, Dec. 8–12, 2003.
[6] S. Borenstein, J. Bushnell, and F. Wolak, "Measuring market inefficiencies in California's wholesale electricity industry," *Amer. Econ. Rev.*, vol. 92, no. 5, pp. 1376–1405, 2002.
[7] N. Fabra, "Tacit collusion in repeated auctions: Uniform versus discriminatory," *J. Ind. Econ.*, vol. 51, no. 3, pp. 271–, Sep. 2003.
[8] E. Dechenaux and D. Kovenock, "Tacit collusion and capacity withholding in repeated uniform price auctions," UFAE and IAE Working Papers 645.05, 2005.
[9] S. S. Oren, "Economic inefficiency of passive transmission rights in congested electricity systems with competitive generation," *Energy J.*, vol. 18, no. 1, pp. 63–83, 1997.
[10] D. W. Bunn and F. S. Oliveira, "Agent-based simulation—An application to the new electricity trading arrangements of England and Wales," *IEEE Trans. Evol. Comput.*, vol. 5, no. 5, pp. 493–503, Oct. 2001.
[11] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, 1992.
[12] G. Xiong, T. Hashiyama, and S. Okuma, "An electricity supplier bidding strategy through Q-Learning," in *Proc. IEEE Power Eng. Soc. Summer Meeting, 2002* , Jul. 21–25, 2002, vol. 3, pp. 1516–1521.
[13] M. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. 11th Int. Conf. Machine Learning*, San Francisco, CA, 1994, pp. 157–163, Morgan Kaufmann.
[14] J. Hu and M. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Mach. Learn. Res.*, pp. 1039–1069, 2003.
[15] Y. Shoham, R. Powers, and T. Grenager, Multi-Agent Reinforcement Learning: Critical Survey, Stanford University, Tech. Rep., 2003.
[16] M. Bowling and M. Veloso, An Analysis of Stochastic Game Theory for Multi-Agent Reinforcement Learning, Carnegie Mellon Univ., Tech. Rep., 2000, CMU-CS-00-165.
[17] T. Krause, G. Andersson, D. Ernst, E. V. Beck, R. Cherkaoui, and A. Germond, "Nash equilibria and reinforcement learning for active decision maker modeling in power markets," in *Proc. 6th IAEE Conf.—Modeling in Energy Economics*, Zurich, Switzerland, Sept. 2004.
[18] T. Krause, G. Andersson, D. Ernst, E. V. Beck, R. Cherkaoui, and A. Germond, "A comparison of Nash equilibria analysis and agent-based modeling for power markets," in *Proc. 15th PSCC*, Liege, Belgium, Aug. 22–26, 2005.
[19] C. D. Aliprantis and S. K. Chakrabarti, *Games and Decision Making*. Oxford, U.K.: Oxford Univ. Press, 2000.
[20] R. Gibbons, "A primer in game theory," *Harvester Wheatsheaf*, 1992.
[21] P. Dayan and C. J. C. H. Watkins, "Reinforcement learning," in *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, R. A. Wilson and F. Keil, Eds. Cambridge, MA: MIT Press.
[22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
[23] C. Ribeiro, "Reinforcement learning agents," *Artif. Intell. Rev.*, vol. 17, pp. 223–250, 2002.
[24] G. Xiong, S. Okuma, and H. Fujita, "Multi-agent experiments on uniform price and pay-as-bid electricity auction markets," in *Proc. IEEE Int. Conf. Electric Utility Deregulation, Restructuring and Power Technologies (DRPT2004), 2004*, Hong Kong, Apr. 2004.
[25] J. Nie and S. Haykin, "A dynamic channel assignment policy through Q-learning," *IEEE Trans. Neural Netw.*, vol. 10, no. 6, pp. 1443–1455, Nov. 1999.
[26] M. Guo, Y. Liu, and J. Malec, "A new Q-learning algorithm based on the metropolis criterion," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 34, no. 5, pp. 2140–2143, Oct. 2004.
[27] N. Metropolis, A. W. Rosenbluth, M. N. Rodenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.
[28] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.
[29] R. Sioshansi and S. Oren, "How good are supply function equilibrium models: An empirical analysis of the ERCOT balancing market," *UCEI Energy Policy and Economics Working Paper Series*, Apr. 2006.
[30] T. Li, M. Shahidehpour, and A. Keyhani, "Market power analysis in electricity markets using supply function equilibrium model," *IMA J. Manage. Math.*, vol. 15, pp. 339–354, 2004.
[31] R. Axelrod, *The Evolution of Cooperation*. New York: Penguin Books, 1990.

Athina C. Tellidou was born in Drama, Greece, in May 1980. She received the Dipl. Electr. Eng. degree from the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, where she is currently pursuing the Ph.D. degree.

Her research interests are in market operation of deregulated power systems.

Anastasios G. Bakirtzis (S'77–M'79–SM'95) was born in Serres, Greece, in February 1956. He received the Dipl. Eng. degree from the Department of Electrical Engineering, National Technical University, Athens, Greece, in 1979 and the M.S.E.E. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, in 1981 and 1984, respectively.

In 1984, he was a Consultant to Southern Company. Since 1986, he has been with the Electrical Engineering Department, Aristotle University of Thessaloniki, Thessaloniki, Greece, where he is currently a Professor. His research interests are in power system operation and control, reliability analysis, and in alternative energy sources.