20 Nov 2015

Strategic Generation Bidding using a Learning Algorithm through Updates of Supply Offer Selection Propensities

Sriram Raghavan

Jhi Young Joo
Missouri University of Science and Technology

Recommended Citation

Raghavan, S., & Joo, J. Y. (2015). Strategic Generation Bidding using a Learning Algorithm through Updates of Supply Offer Selection Propensities. *2015 North American Power Symposium, NAPS 2015* Institute of Electrical and Electronics Engineers.
The definitive version is available at https://doi.org/10.1109/NAPS.2015.7335223

Strategic Generation Bidding Using A Learning Algorithm Through Updates Of Supply Offer Selection Propensities

Sriram Raghavan, and Jhi-Young Joo, *Member*, *IEEE*
Department of Electrical and Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri, USA
Email: {sr257, joojh}@mst.edu

***Abstract*— This paper discusses a novel bidding strategy of a generation company (genco) in an hourly day-ahead market. In the proposed method, a genco learns the returns of supply offers and adapts its strategy accordingly, based on the Variant Roth-Erev (VRE) reinforcement learning algorithm. Every supply offer submitted to the market receives a profit at the end of each day, and is strategically updated for the next day based on this profit. The novelty of our proposed method is that every supply offer has a propensity (an inclination or a tendency) to be selected associated with it. The propensity is updated as a percentage relative to every other supply offer's propensity based on the profit received. The DC optimal power flow problem solved by the system operator is also improved by including the physical inter-temporal constraints such as the generator ramp rates, in addition to the supply offers. Simulations on a 5-bus test system show that a genco learns to strategically bid in the market using the relative percentage propensity update technique. As a result, without any market regulations, the locational marginal prices increased by 29% on average.**

Keywords—strategic bidding, electricity market, reinforcement learning, generator bidding, DC optimal power flow

I. Introduction

Generation of electric energy is scheduled to meet forecasted demand through a day-ahead electricity market in most deregulated industries. This particular forward market operates a day in advance of the actual physical delivery of power. The decision on power generation for the next day in the market is a result of a two-sided auction where gencos (power producers; supply agents) and load serving entities (demand agents) submit a set of price-quantity bids [1]. As such, a genco needs to decide its supply bids in order to secure the maximum possible profits for itself.

A considerable amount of work has been done in the past regarding strategic bidding by gencos to improve their profits. The Agent Based Modeling of Electricity Systems (AMES) Wholesale Power Market Test Bed is an agent-based computational laboratory for studying the dynamic performance of restructured wholesale power markets [3]. AMES models gencos with reinforcement learning capabilities interacting over time with electric power buyers (load-serving entities) in a wholesale power market. [4] applies optimal control to study generator bidding in an oligopolistic electricity market. Game theory has been used to describe a simple method to derive strategic equilibrium solutions for a single genco bidding in electricity markets in [5]. [6] discusses the effect of generators' strategic behavior on individual's payoffs and market efficiency by studying the generator bidding using stochastic optimal control.

In this paper, we use reinforcement learning to decide supply offers (marginal cost coefficients and maximum production levels) that the gencos bid in the day-ahead market. The gencos use learning algorithms that adjust bid supply offers to produce power in the electricity market. The supply offers can be improvised or adjusted for the next day when the gencos analyze the profits that they make at the end of each day's market clearance. This helps them secure higher profits with time in a wholesale electricity market. The main objective of this paper is to investigate whether the gencos can employ strategic supply offers given inter-temporal physical constraints of system components. This is done by modifying the VRE reinforcement learning algorithm. A five-bus transmission grid test system [2] is used to show the results. We also show that all generators learn over time to implicitly report higher than true marginal costs, thus considerably raising the value of the locational marginal price of the system.

II. Formulations For System Operation

A. The problem of a genco

In this model, without loss of generality, we assume that each genco owns a single generator. The cost function of a generator has a variable and fixed cost of generation. For simplicity, we do not consider no-load, startup or shut down costs. Each generator's lower and upper production limits are denoted by $LCap_i$ and $UCap_i$ in MWs, that define the feasible production interval for its hourly real-power production level P_{Gi} (in MWs) where *i* represents each genco.

For each generator *i*,

$$LCap_i \leq P_{G_i} \leq UCap_i \qquad (1)$$

Genco i's total cost function giving its total costs of production per hour for each P_{Gi} takes the form

$$TC_i(P_{Gi}) = FC_i + a_i.P_{Gi} + b_i.P^2{}_{Gi} \quad (2)$$

where a_i (\$/MWh), b_i (\$/MWh^2), and FC_i (\$/h) are given constants. The marginal cost function for Generator i is given by

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

The scheduling of generators on day D is a result of the market clearance by the ISO computed on the previous day D-1. At the beginning of each day D-1, Genco i submits a supply offer $s_i^R(D\text{-}1)$, a function of each day (for simplicity, gencos submit a single bid for each of the 24 hours) to the system operator for use in each hour H of the day-ahead market for day D.
This supply offer consists of a reported marginal cost function

$$MC_i^R(P_{Gi}) = a_i^R + 2 \cdot b_i^R \cdot P_{Gi} \quad (4)$$

defined over a reported feasible production interval

$$LCap_i^R \leq P_{Gi} \leq UCap_i^R \quad (5)$$

The reported cost coefficients a_i^R and b_i^R differs from the generator i's true cost coefficients a_i and b_i after the process of strategic learning.

Let generator i, located at bus k, report a supply offer $s_i^R(D)$, for the day D (along with the market participants) on the previous day $D-1$. Let LMP_k be the locational marginal price at bus k calculated by the system operator for hour H on day D, and P_{Gi}, the real power that Genco i has been cleared to produce in hour H of day D. The profit of Genco i from the day D settlement of this financial contract for hour H of day D is given by

$$\text{Profit}(P_{Gi}) = \text{LMP}_k.P_{Gi} - TC_i(P_{Gi}) \quad (6)$$

B. The problem of a Load Serving Entities (LSE)

The LSEs purchase energy in the day-ahead market each day in order to meet the demand (load). We assume that the LSEs do not involve in production or sale of energy in the wholesale market and thus purchase energy only from generators and not from each other. A daily load profile is submitted into the day-ahead market as demand bids without any strategic learning for day D at the beginning of day $D-1$ for each of the 24 successive hours. For simplicity, the hourly system load is assumed to be the same for all days.

C. The problem of the Independent System Operator (ISO)

The activities of the ISO during a day D is shown in Fig. 1. The ISO in our model during each day D determines a schedule of optimal power commitments for each hour of the day-ahead market conditional to the supply offers submitted by gencos, demand bids submitted by LSEs, branch flow limits, and nodal balance constraints ensuring the total supply meets the total demand.

The resulting optimization problem is known as a bid-based DC optimal power flow (OPF). This is a convex quadratic programming problem when the bids of gencos are linear to their respective supply output.

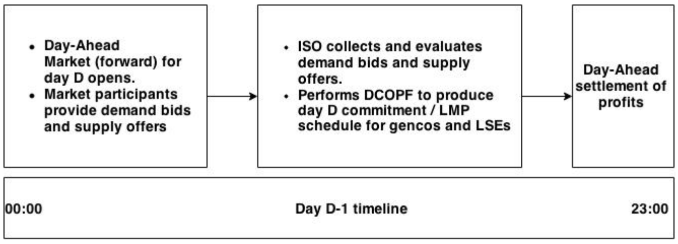

Fig. 1. Activities of the ISO on day D-1 for day-ahead market

The objective of the problem as stated in [7] can be formulated as follows:

$$\min \sum_{i=1}^{m} [FC_i + a_i P_{Gi} + b_i P_{Gi}{}^2]$$

subject to:

$$LCap(P_{Gi}) \leq P_{Gi} \leq UCap(P_{Gi}) \quad (6)$$

$$P_{GK} - P_{DK} = B\theta_K \quad (7)$$

$$-P_{ik\max} \leq \frac{1}{X_{ik}}[\theta_i - \theta_k] \leq P_{ik\max} \quad (8)$$

Equation (6) represents the generator output constraint. Equation (7) represents the active power balance for the k$^{\text{th}}$ bus of transmission system (nodal power balance constraint) where P_{GK} and P_{DK} are active power generation and demand at bus k, B is the susceptance of the system, θ_k is the voltage angle at bus k. In (7), $P_{ik\max}$ is the maximum power that can flow in the branch i-k, X_{ik} is the line reactance of branch i-k and θ_i and θ_k are the respective voltage angle setting. We use MATPOWER [9] to simulate this problem.

III. The Variant Roth Erev Learning Algorithm For Gencos' Bids

The objective of a genco is to maximize its profits at the end of each day. Submitting supply offers with expensive prices can return high profits if accepted. However, the likelihood of expensive bids being accepted is lower, since the other suppliers may submit lower-price bids.

Therefore, a genco has to strategically bid in supply offers that are low enough to be cleared by the system operator as well as sufficient enough to make profits for itself. This can be made possible if the gencos can analyze their profits at the end of each day, learn over time on how to improve their bids based on their rewards (profits) and achieve better profits with time.

In [3], gencos adaptively select their supply offers based on their past profit outcomes using stochastic reinforcement learning algorithm developed by Roth-Erev, referred to as the Variant Roth Erev (VRE) learning algorithm. We adopt the same methodology, but modify the algorithm to make it a lot easier in terms of calculation and understanding. The learning processes of the gencos are shown in Fig.2.

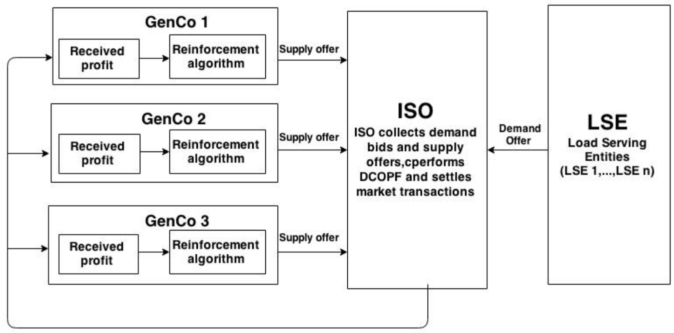

Fig. 2. Learning process of gencos for strategic bidding

At the starting of the initial day D=1, Genco i chooses a supply offer from its action domain AD_i to report to the ISO for the day-ahead market in day D+1. The action domain AD_i is a collection of supply offers in a matrix form. Numerous supply offers are generated through mathematical formulations to provide a flexibility for the gencos to choose and submit from. The supply offer is picked from the matrix based on the VRE reinforcement learning algorithm and is submitted to the system operator for the day-ahead market. The building of the action domain matrix AD*i* is explained in [1].

Every supply offer in AD_i is assigned with a propensity. The propensity of a supply offer m indicates the genco's profit expectation at the end of each day when the supply offer m is submitted. The initial propensity of Genco i to choose a supply offer m from AD_i is given by q_{im}^0. We fix a constant value q_i^0 such that $q_{im}^0 = q_i^0$ for all supply offers in $m \in AD_i$ to allow random selection of supply offer on day D=1.

Let the propensity of generator i choose supply offer $m \in AD_i$ be $q_{im}(D)$ on day D. The choice probability that Genco i uses to select a supply offer for day D is then constructed from these propensities as follows:

$$P_{in}(D) = \frac{\exp(q_{im}(D)/C_i)}{\sum_{j=1}^{M_i} \exp(q_{ij}(D)/C_i)}, \quad m \in AD_i \tag{9}$$

where the propensity is updated every day with respect to the profit as shown below

$$q_{im}(D+1) = [1-r_i]q_{im}(D) + \text{Response}_{im}(D)$$

where, (10)

$$\text{Response}_{im}(D) = \begin{cases} [1-e_i]\text{Profit}_{im'}(D) & \text{if } m = m' \\ \dfrac{e_i q_{im}(D)}{M_i - 1} & \text{if } m \neq m', \end{cases}$$

where m≠m□ implies that there are more than two supply offers.

The parameters used in the reinforcement learning algorithm are explained in [8] and are summarized as follows:

- Boltzmann cooling parameter (C_i): The Boltzmann cooling factor calculates the action choice probability, given their respective propensity. An appropriate selection of the cooling factor helps in determining the probability when the difference between propensities of two supply offers is very small.
- Experimentation (e_i): The experimentation factor (e_i) controls the frequency of selecting new supply offers from the action domain AD_i. A higher value of experimentation factor gives the genco flexibility to experiment new supply offers more frequently and vice versa. The purpose of an experimentation factor is to prevent a genco from choosing a single supply offer at a very early stage.
- Recency (r_i): The recency factor is used to minimize the selection of a supply offer that was selected for a longer timeline before. This enables a genco to ignore profits obtained by the supply offer before and helps them "learn" the current scenario of market and submit bids accordingly.

Once the values for the initial propensity value q_i^0, the experimentation parameter e_i are designed given specific requirements, Genco i starts learning to strategically bid in the electricity market.

IV. Relative Percentage Update Of Supply Offer's Propensity

According to the VRE reinforcement learning algorithm, the initial propensity of every supply offer in the action domain matrix AD_i for Genco i is chosen to be a finite real number. When the initial propensity chosen is too low compared to the expected profit from the genco's action, the genco fixates on a particular supply offer at a very early stage. This reduces the probability of choosing other supply offers to an extremely small value. When the initial propensity is too high, the genco will cycle through a lot of supply offers. To avoid the complexity, we choose the propensity and profits of the supply offers to be in percentages with respect to every other supply offer and update them based on the profit incurred by the chosen supply offer.

Fig. 3 shows the configuration with five gencos (each with one generator) and three across a 5-bus transmission grid [2]. The generator cost functions and ramp rates are given in Table I. The ramp rates (% capacity per minute) of generators G1, G2, G3, G4 and G5 are 40, 4, 3, 3 and 1 respectively. This five-bus transmission grid configuration [2] is used extensively in ISO New England and PJM training manuals to solve for bid-based DC optimal power flow (OPF) solutions using supply offers by gencos and demand bids by LSEs. This problem is solved subject to branch flow limits and nodal supply-demand balance at each bus [1]. The grid branch and load input data for the dynamic five-bus test case are taken from [1].

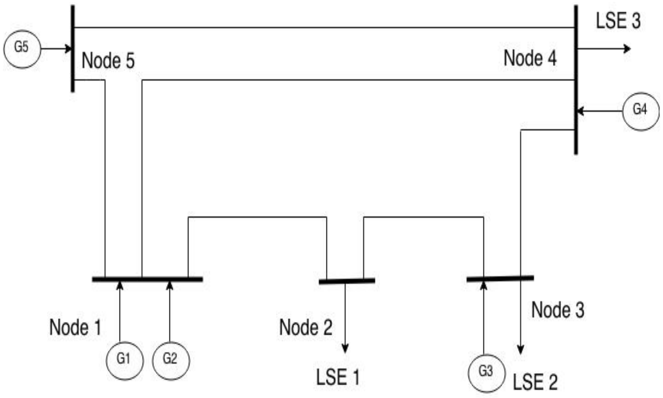

Fig. 3. John Lally 5-bus system [2]

TABLE I. Gencos' cost functions

Gen ID	Bus	FC_i ($)	b_i ($/MWh)	a_i ($/MWh^2)	Cap (MW)
1	1	832.75	48.58	0.0072	110
2	1	100.00	0.001	0.00001	100
3	3	665.10	11.85	0.0049	520
4	4	382.23	12.38	0.0083	200
5	5	395.37	4.42	0.0002	600

To validate the proposed approach, we solve an example for one of the gencos, Genco 4, by updating its propensity for the day D based on the profit incurred at the end of day $D-1$. At the start of Day 1, all the initial propensity $q_{4m}(1)$ of the supply offers are equal. For simplicity, let m = [1, 2…, 5], i.e. the action domain AD_4 contains five supply offers to be chosen from and submitted. Table II below represents the supply offers, their initial propensities and its respective action choice probabilities for the example. As mentioned before, all the propensities of the supply offer are written as a relative percentage of each other.

TABLE II. Genco 4's supply offer selections on Day 1

Supply offer m	Supply bids b_i ($/MWh)	Supply bids a_i ($/MWh^2)	Propensity $q_{4m}(1)$ (percentage)	Action choice probability $P_{4m}(1)$
1	12.38	0.0083	20	0.2
2	12.38	0.0346	20	0.2
3	13.76	0.0147	20	0.2
4	13.76	0.0311	20	0.2
5	15.48	0.0104	20	0.2

Since all the action choice probabilities are equal by the setup, supply offer 1 is randomly chosen and reported to the ISO on Day 1. For simplicity, supply offer 1 is assumed to be the true marginal cost of Genco 4. The system price and quantities are settled at the end of the day by the ISO (using equations (2) and (6)). The profit for each genco is calculated according to the cleared market price and supply quantity. BPCG (Bid Production Cost Guarantee) is a guarantee provided by the ISO to the Gencos where in a unit will not incur a net loss, if dispatched. The BPCG payment made to Gencos is settled on a daily basis. In order for a Genco to receive a BPCG payment, the net sum of loss/profit incurred every hour must result in a net loss. This paper assumes no BPCG as the objective of this model is to see if Gencos can learn from their results (loss/profit). The BPCG could be implemented in the future work to see its significance in the learning of Gencos.

Total revenue of Genco 4 on Day 1 when supply offer 1 is chosen = $4576.371
Total cost of production on Day 1 = $13107.528
Profit on Day 1 = −$8531.157

Clearly, Genco 4 has incurred a loss on Day 1 when supply offer 1 is chosen. Hence, the propensity and action choice probability of offer 1 being chosen again on Day 2 should be lowered. Thus, based on the loss (negative profit), the recency and the experimentation factor, the propensities are updated for Day 2. Another important reason why gencos must learn to strategically bid in the electricity market is seen in the above example. Genco 4 in the example submitted the true marginal cost as their supply offer and incurred loss on Day 1. Genco 4 has to analyze this loss, learn what offers incur losses, and achieve better profits instead of submitting their true marginal cost.

The propensities and action choice probabilities (using equation (9) and (10)) are updated and are converted to relative percentages and are shown in Table III below. Clearly, the action choice probability of supply offer 1 dropped to 0.1979 from 0.2 because its percentage propensity of supply offer 1 decreased from 20% to 18.96%.

TABLE III. Genco 4's supply offer selections on Day 2

Supply offer (m)	Supply bids b_i ($/MWh)	Supply bids a_i ($/MWh^2)	Propensity $q_{4m}(2)$ (percentage)	Action choice probability $P_{4m}(2)$
1	12.38	0.0083	18.96	0.1979
2	12.38	0.0346	20.24	0.2005
3	13.76	0.0147	20.24	0.2005
4	13.76	0.0311	20.24	0.2005
5	15.48	0.0104	20.24	0.2005

The entire procedure of the model is listed below:

Step 1: On Day 1, the gencos choose a supply offer randomly from their respective action domain AD_i and submit them to the ISO. The LSE submit their demand bids as well.

Step 2: The ISO solves the DC OPF with the given supply offers, demand bids, branch thermal limit constraints and nodal constraints. The LMPs are calculated as a result of the DC OPF.

Step 3: Each genco calculates its profit from the market settlement at the end of day.

Step 4: The propensity and choice probability for the supply offers are now updated in the action domain AD_i of Genco i using the reinforcement learning algorithm.

Step 5: The genco chooses the best supply offer using the action choice probability for the next day. ISO clears the market with the supply offers from the gencos and demand bids from the LSEs, going back to Step 2. The procedure is repeated for every day.

V. Simulations and Results

For the above system [2], we run simulations for 30 days, with each genco containing six supply offers to choose and submit from. Even though the simulation is done for 30 days, it should be noted that the hourly system load profile remains the same over these days. This assumption is to evaluate how gencos adapt their strategies for the same load profile, and should be relaxed in future work. For simplicity, we also assume that the gencos submit a single supply offer for one day (same offer for all 24 hours) and each genco has only one generator. The gencos' cost functions are shown in Table II.

The hourly ramp rates of the generators are included when calculating the generator dispatch by the Independent System Operator (ISO) using DCOPF. This means that each genco submits its ramp rates to the ISO along with its supply offer for the day. The bid-based DC OPF solved by the system operator is simulated using MATPOWER [9]. The LMPs at different buses on Day 1 are shown in Fig. 4. The profits that are settled at the end of Day 1 by the system operator are updated using the proposed alternative VRE reinforcement algorithm to submit new and better bids on Day 2 and so on.

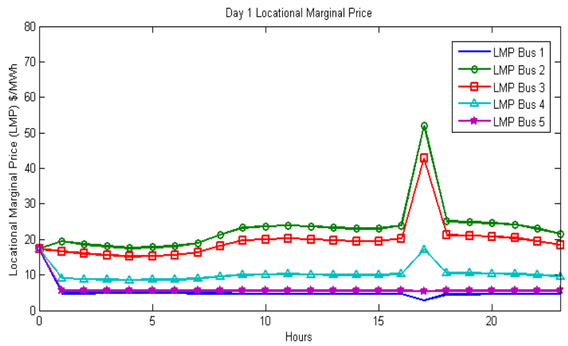

Fig. 4. LMPs on Day 1 at different buses

In this simulation, the gencos report strategic supply offers and their true production limits. The gencos learn over time on what prices to offer in order to increase their net profits. For example, the peak load at Hour 17 (load input data from [1]) cannot be met even with the combined capacity of the smallest three Genco 1, 2, and 3. In other words, this peak load cannot be met without Genco 4 and 5. Hence, if these gencos have the information on the cleared bids and the profits secured at the end of day repeatedly, they can exercise market power if their highest reported supply offer within their action domains was still rewarded.

At Hour 1, the LMPs of the system are the same. This is due to the ramp constraint of Genco 5. In the first hour, Genco 5 can produce only 350 MW due to which there is no congestion in the system. Hence, Genco 4 has to produce power to meet the load at Hour 1. After the first hour, Genco 5's production level increases and hence there is congestion in the system due to which the LMPs differ. Figs. 5 and 6 show the supply offers of Genco 4 and 5 during the simulation as a result of their learning. There is a large difference between the true marginal costs and the supply offers. It is obvious that when all the gencos submit higher bids, the LMPs at different buses will increase. This is shown in Fig. 7. The LMP values at different buses from Day 1 to Day 30 increase by an average of 29%.

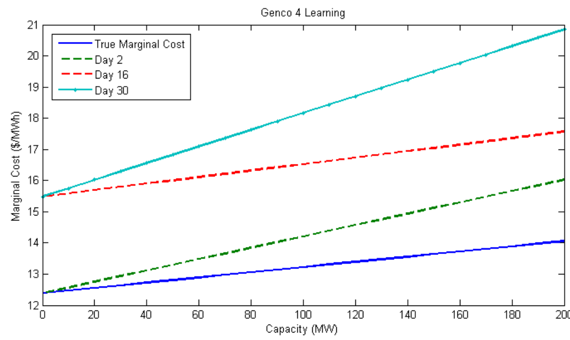

Fig. 5. Genco 4 learning

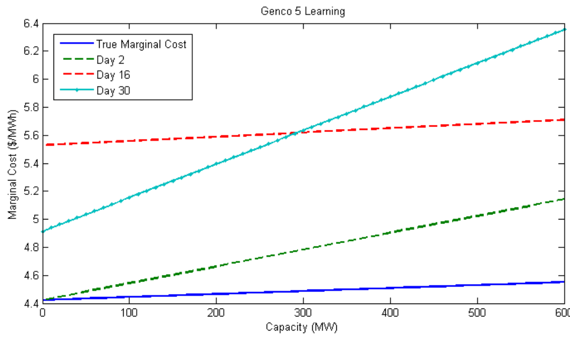

Fig. 6. Genco 5 learning

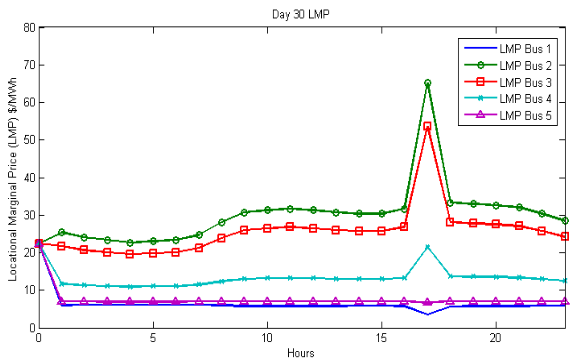

Fig. 7. LMPs on Day 30 at different buses

VI. CONCLUSION

A bidding strategy of a genco in an hourly day-ahead market using VRE learning algorithm was proposed. Simulation on a test system [2] shows that the gencos learn to increase their profits over time. This can be verified with the graph showing the comparison of their true cost function and their reported supply offer on different days. An average increase of 29% is seen in the LMPs in the 5-bus test system case.

This paper implemented a bidding strategy for an hourly day-head market where a single supply offer is submitted for the whole day. We plan to find the effect of the learning algorithm on submitting hourly varying supply offers for the day ahead market with the inter-temporal constraints. More rigorous analysis on the effects of ramp rates on hourly bids and on LMPs will be conducted. Further research is also needed to explore how the strategies of a genco should adapt to ever-evolving daily system load profiles.

REFERENCES

[1] J. Sun and L. Tesfatsion, "Dynamic testing of wholesale power market designs: An open-source agent-based framework," Computational Economics, Vol.30, pp. 291–327, 2007.

[2] J. Lally, "Financial transmission rights: Auction example," ISO New England M-06, Tech. Rep., January 2002

[3] Hongyan Li; Tesfatsion, L., "The AMES wholesale power market test bed: A computational laboratory for research, teaching, and training," *Power & Energy Society General Meeting, 2009. PES '09. IEEE* , vol., no., pp.1,8, 26-30 July 2009

[4] Youfei Liu; Wu, F.F., "Generator bidding in oligopolistic electricity markets using optimal control: fundamentals and application," *Power Systems, IEEE Transactions on* , vol.21, no.3, pp.1050,1061, Aug. 2006

[5] Morinec, A.G.; Villaseca, F.Eugenio, "Optimal generator bidding strategies for power and ancillary services using game theory," *Power Symposium, 2008. NAPS '08. 40th North American* , vol., no., pp.1,8, 28-30 Sept. 2008

[6] Youfei Liu; Wu, F.F., "Prisoner Dilemma: Generator Strategic Bidding in Electricity Markets," *Automatic Control, IEEE Transactions on*, vol.52, no.6, pp.1143,1149, June 2007

[7] Francisco M. Gonzalez-Longatt, José Luis Rueda: *Power Factory Applications for Power System Analysis*. Springer 2014

[8] Mridul Pentapalli. 2008. A comparative study of Roth-Erev and modified Roth-Erev reinforcement learning algorithms for uniform-price double auctions. Master's Thesis, Iowa State University, Ames, Iowa

[9] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education," *Power Systems, IEEE Transactions on*, vol. 26, no. 1, pp. 12-19, Feb. 2011