Adaptive Bidding for Electricity Wholesale Markets in a Smart Grid

Jurica Babic and Vedran Podobnik

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{jurica.babic,vedran.podobnik}@fer.hr
http://agents.tel.fer.hr

Abstract. One of challenges for researchers in smart grids is to find mechanisms for putting orders on the electricity wholesale market. This paper tackles this problem by proposing adaptive bidding mechanism for trading in the wholesale market. The research challenge lies in the fact that wholesale players simultaneously trade on 24 different wholesale markets determined by the moment of electricity delivery which ranges from 1 to 24 hours ahead. Namely, the variant Roth-Erev reinforcement learning algorithm is used to coordinate wholesale bidding across different markets by choosing among four implemented wholesale strategies. The Power Trading Agent Competition is used to evaluate the performance of different implementations of the adaptive bidding mechanism as well as to benchmark adaptive bidding approach against single strategy approach.

Keywords: smart grids, software agents, electricity wholesale market, reinforcement learning, Power Trading Agent Competition

1 Introduction

Liberalization and decentralization of electricity markets has resulted in major changes of their structure and dynamics, thus creating a regulated and competitive market environment. To enable further improvements, most of traditional power grids are introducing novel solutions based on information and communications technology (ICT), progressively transforming into systems called smart grids that enable more efficient energy usage, better communication between entities on the market as well as real-time balancing of energy supply and demand. One of challenges for researchers in smart grids is to find mechanisms for putting orders on the electricity wholesale market. This paper addresses raised challenge by proposing an adaptive bidding mechanism for putting orders on the electricity wholesale market. The research challenge is even more complex having in mind that we consider market design where wholesale players simultaneously trade on 24 different wholesale markets determined by the moment of electricity delivery which ranges from 1 to 24 hours ahead.

In particular, the wholesale market we consider represents an energy market where agents (i.e., retail brokers) may engage in trading along with dedicated

Babic, Jurica; Podobnik, Vedran.
Adaptive Bidding for Electricity Wholesale Markets in a Smart Grid.
Proceedings of the Workshop on Agent-Mediated Electronic Commerce and Trading Agent Design and Analysis (AMEC/TADA 2014) @ AAMAS 2014 / Ceppi, S. ; David, E. ; Robu, V., Shehory O. ; Vetsikas, I.A., editor(s).
Paris, France: International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2014. 1-14.

wholesale players such as large energy producers (i.e., generation companies or GenCos) which provide necessary bulk energy and market liquidity. Brokers are able to trade for future delivery, i.e., between 1 and 24 hours in the future where a slot (i.e., enabled timeslot) for every hour in the future is represented as a different market. The wholesale market works as a periodic double auction, meaning it enables almost real time trading by clearing the standing order books once every simulated hour. The clearing process considers the order book of the enabled timeslot, whose orders are used to construct supply and demand curves and consequently to determine the clearing price for the enabled timeslot. Orders are sent by the brokers and must include the information about offered price and the amount of energy. Orders with positive amount of energy (i.e., order issuer wants to buy energy) are considered *bids*, and those with negative amount of energy (i.e., order issuer wants to sell energy) are considered *asks*. The clearing process sorts bid orders from the highest to lowest price and ask orders from lowest to highest price and determines the clearing price at the intersection of the supply and demand curves. In case there is no intersection between curves, the clearing price is set at the mean of the lowest bid price and the highest ask price. Trades are made for all asks with prices below the clearing price and for all bids with prices above the clearing price.

This paper proposes a mechanism based on the variant Roth-Erev reinforcement learning algorithm for coordination of wholesale bidding across different markets by choosing among four different wholesale strategies that are implemented in the software agent. We use Power Trading Agent Competition (TAC) [4] to evaluate the proposed adaptive bidding strategy system. Power TAC is a competitive economic simulation of the smart grid that aims to provide an insight into the structure and operation of electricity markets in a smart grid environment [3]. In the Power TAC simulation competitors are brokers that provide electricity services to retail customers using tariff offerings, while managing their customer portfolio by trading in a wholesale market.

The approach based on reinforcement learning was already studied by Power TAC researchers. However, it was used for other parts of Power TAC system such as retail market operations optimization [6] or customer modelling [9]. This paper presents an experiment to evaluate the performance of different implementations of the adaptive bidding mechanism as well. The evaluation process is designed based on previous papers on the Power TAC benchmarking [2][5]. The paper is organized as follows. Section 2 presents a mechanism for wholesale trading based on Adaptive Bidding System (ABS). Afterwards, Section 3 explains the experiment setup and methodology as well as it discusses the preliminary results regarding ABS. Section 4 concludes the paper with the ideas for the future work.

2 Wholesale Trading Based on Adaptive Bidding System

The Adaptive Bidding System (ABS) is a trading module implemented within the CrocodileAgent broker [1], a software agent developed by the University of Zagreb for the annual Power TAC competition (www.powertac.org). The ABS is used for trading in the wholesale market. Fig.1 depicts a high-level architecture of the proposed solution for trading in electricity wholesale markets. The *CrocodileAgent* broker interacts with the Power TAC environment by using:

- *Retail module* to interact with the *retail market*. This is the module responsible for managing customer portfolio (e.g., new tariff offerings) and predicting the expected energy load from its customers;
- *Context repository* for storing all relevant data from the environment such as the *retail context* (e.g., public and private tariff offerings, retail transactions), *game context* (e.g., bank transactions, weather information) and *wholesale context* (e.g., cleared trades, order books);
- *Wholesale module* with ABS for trading energy on the *wholesale market*.

Essentially, the ABS may be conceived as a black box with two *inputs*: (i) the expected amount of needed energy to be acquired, and (ii) historical data on wholesale transactions and balancing transactions; and one *output*: a generated order (i.e., a bid or an ask) which is sent to the wholesale market.

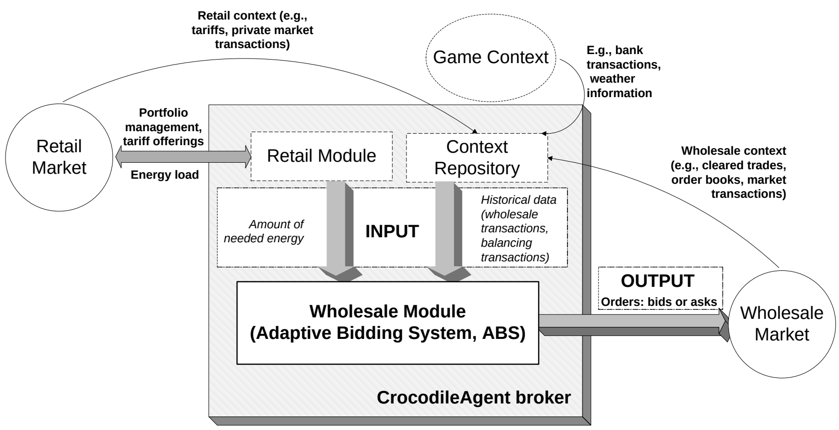

Fig. 1. The CrocodileAgent broker interacts with the Power TAC environment.

Fig. 2 shows more details on the internal structure of the wholesale module. Inside the black box there is the *bucket master* which performs following tasks:

- chooses the specific *learning bucket* depending on the time proximity;
- delegates the job to the selected learning bucket and most importantly;
- performs learning via embedded *learning algorithm*.

The *time proximity* is defined as the distance between the current timeslot and the target timeslot (i.e., the timeslot for which the order is made). A *learning bucket* is an artefact responsible for the range of proximities and it has (i) strategy portfolio with four distinct wholesale strategies (hereinafter: strategies) and (ii) knowledge repository with knowledge used in the learning process (i.e., propensities and probabilities for the strategies). Within the ABS there can be 1, 2, 3, 4, 6, 8, 12 or 24 learning buckets, chosen in the initial setup. Finally, the *performance reporter* receives the performance data about the ABS and forms a log which can be processed and analysed.

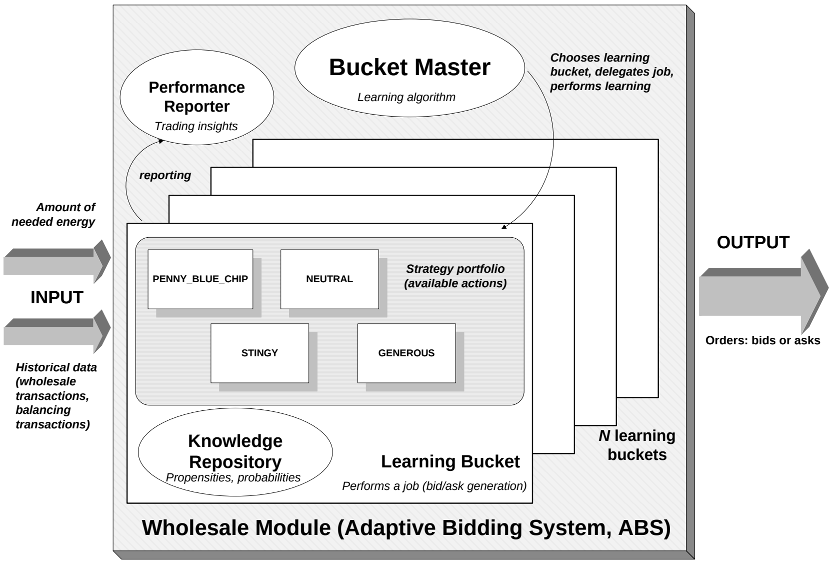

Fig. 2. Building blocks of the wholesale module.

2.1 Learning Buckets

As introduced in the previous subsection, *learning buckets* are responsible for the actual order formation and for dispatching the order onto the wholesale market. The main idea behind learning buckets is the fact that, under most scenarios, a reasonable broker might want to play different strategies depending on the remaining chances (i.e., timeslot proximities). Adaptable broker behaviour in trading scenarios is indeed achievable with the use of learning algorithms. However, by introducing learning buckets in the system the CrocodileAgent broker is also able to perform trading depending on the timeslot proximity and to track the learning experience within the corresponding timeframe supported by a learning bucket. Therefore, the proposed ABS performs two-dimensional optimization of a bidding behaviour by taking into account both strategy execution outcomes and the timeframe when the strategies where performed.

Essentially, a learning bucket B is defined by its starting index *bucketIndex*, size *bucketSize* and the set of wholesale strategies (i.e., strategy portfolio SP). Fig. 3 shows the example of the ABS configuration where there are three learning

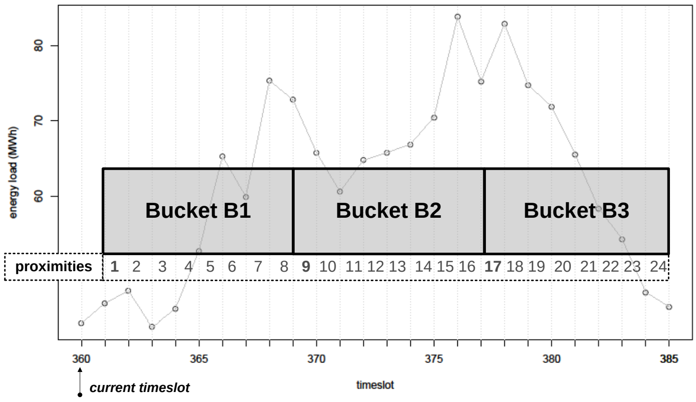

Fig. 3. Example configuration with three learning buckets *B1, B2, B3* where each of them is responsible for a particular timeframe.

buckets: *B1*, *B2* and *B3*. Since there are total of 24 proximity timeslots, the *bucketSize* of each is eight. Each of them is responsible for only a portion of timeslot proximities. Learning buckets *B1*, *B2* and *B3* have $startIndex_{B1} = 1$, $startIndex_{B2} = 9$, $startIndex_{B3} = 17$, respectively. The figure also shows how timeslot proximities are mapped to an ordered set of timeslots. Since the observation point is set at the timeslot 360, the last-minute trading timeslot proximity (i.e., proximity equals to one) is for timeslot 361 and it has a value of 1. Similary, the furthest timeslot opened for trading is for the timeslot 385 and the corresponding timeslot proximity equals to 24. Finally, the graph line shows the energy load for each of timeslot proximities which emerges from customers (i.e., producers and consumers) within the broker's portfolio and it is one of inputs for a wholesale order generation.

2.2 Wholesale Strategy Portfolio

In the context of ABS, the strategy is a policy for making the final order to be placed on the wholesale market by using the baseline price $price_{base}(t)$ and baseline energy $energy_{base}(t)$.

The baseline price $price_{base}(t)$ for the timeslot t is retrieved from the context repository and it is modeled to the weighted daily mean price by (i) determining the hour of the day of the input timeslot t; and by (ii) considering trades of surrounding hours. For example, if the context repository determines that the timeslot t represents 10:00AM, the $price_{base}(t)$ will consider recent trades which happened at 9:00AM, 10:00AM and 11:00AM. In attempt to capture the volatile nature of daily price trends, the context repository will only remember the most recent trades for the hour of the day, specified by the size of repository's sliding window.

The baseline energy $energy_{base}(t)$ for the timeslot t is retrieved from the retail module which calculates the expected energy load for customers from the broker's customer portfolio at the given timeslot t. The retail module updates the energy usage data for each customer type (e.g., households or office complexes) by exponential smoothing.

There are four wholesale strategies available within the CrocodileAgent ABS which cover some of typical behaviour patterns on the wholesale market.

Neutral Strategy
: As the name implies, the *Neutral* strategy is a risk-averse strategy which attempts to buy or sell the same amount of energy as the baseline energy $energy_{base}$ at the price which closely follows the baseline price.

Penny Blue Chip Strategy
: The *Penny Blue Chip strategy* (or *Penny* strategy in short) is an exploitation strategy which buys cheap (i.e., under the baseline price $price_{base}$) and more than needed energy. In case the broker has an excess of energy, the strategy will attempt to sell all capacities at the price significantly higher then the baseline price (i.e., the expected mean price).

Stingy Strategy
: The *Stingy strategy* is a more extreme version of the *Penny* strategy because it attempts to buy energy at the price lower than the price offered by the *Penny* strategy and sell the excess energy at the price higher than the price offered by the *Penny* strategy. This strategy promises lucrative payoffs but at the higher risk of not making a trade than other strategies within the strategy portfolio.

Generous Strategy
: As the name implies, *Generous* strategy is a last-resort strategy which buys at the price higher than the price offered by *Penny* strategy and sells at the price lower than the baseline price. This strategy may be used when the agent is running out of options (i.e., agent did not secure a trade) in the final timeslot proximities.

Table 1. Multiplication factors of baseline price and energy for selling orders.

	Selling policy		Buying policy	
Strategy	Price	Energy	Price	Energy
Penny	1.6 ± 0.1	1.0	0.8 ± 0.1	1.6 ± 0.1
Neutral	1.0 ± 0.1	1.0	1.0 ± 0.1	1.0
Stingy	1.9 ± 0.1	1.0	0.5 ± 0.1	1.1 ± 0.1
Generous	0.6 ± 0.1	1.0	2.5 ± 0.2	1.0

Strategies also check whether there was a price escalation by checking the outcome of the last corresponding order. If the last order failed to trade, the strategy will check whether the baseline price is better (i.e., more likely to trade) than the last order. If not, the baseline price will be adjusted to be closer towards the last trade's price. Table 1 provides details on how strategies modify the baseline price and baseline energy and consequently generate final orders for placing on the wholesale market. Although the ABS contains only four strategies, its modular design and extendibility enables the support of fine-grained bidding behaviour such as those proposed by authors in [10].

2.3 Variant Roth-Erev Reinforcement Learning Algorithm

The variant Roth-Erev reinforcement learning algorithm (VRE RL algorithm) [8] is the modification of the original Roth-Erev algorithm which was proposed by Albin Roth and Ido Erev [7]. The main idea of the algorithm is to mimic the way human subjects behave under the given scenario. The VRE RL algorithm improves the original Roth-Erev algorithm by (i) enabling the agent to learn in case the action taken has zero payoff, and (ii) keeping probabilities of actions with nonnegative values even in case rewards can take on negative values.

In the context of the proposed ABS, the *action* is considered as the *wholesale strategy*, i.e., a policy for preparing an order to be placed on the wholesale market. The VRE RL algorithm assigns equal initial propensity value $q(0)$ to each of the N actions of the learning agent for $t = 0$. The action propensities at $t+1$ based on the action j, chosen at time t are given with the following expressions:

$$q_j(t+1) = \begin{cases} (1-r) \times q_j(t) + \pi_k(t) \times (1-e), & \text{if } j = \text{k} \\ (1-r) \times q_j(t) + q_j(t) \times \dfrac{e}{N-1}, & \text{if } j \neq \text{k} \end{cases} \tag{1}$$

where e is the experimentation parameter, r is the recency parameter and k is the chosen action.

The *experimentation factor* e is used to comprise the influence of learning exploration and exploitation. A higher experimentation parameter increases propensity values for actions not chosen at time t. Consequently, probabilities of actions not chosen will also increase. This parameter also works as a countermeasure for premature fixation on the chosen action in cases where the chosen action gets a positive reward at the early stage of the experiment.

The *recency factor* r controls degree by which the agent will neglect the past experiences obtained over the experiment. A higher recency factor will cause greater discount on propensity values of actions accumulated in the past. This feature is important in highly dynamic systems where old rewards may become irrelevant.

The propensity value $q_j(t)$ for action j at time t is mapped to choice probability $p_j(t)$ with the use of Gibbs-Boltzmann distribution:

$$p_j(t) = \frac{e^{\frac{q_j(t)}{\tau}}}{\sum_{i=o}^{N-1} e^{\frac{q_j(t)}{\tau}}} \tag{2}$$

where τ is the Boltzmann cooling parameter, used to control mapping of action propensities to probabilities (e.g., for $\tau \to \infty$ action probabilities will be uniform $1/N$).

The proposed ABS uses VRE algorithm to learn what action (i.e., wholesale strategy) is the most appropriate to play on the wholesale market by using the reward function $\pi_k(t)$.

2.4 Reward Function

The wholesale strategy j chosen at time t is rewarded with the reward function $\pi_k(t)$ defined as:

$$\pi_k(t) = \rho \times \mu_b \times \omega(t) + \xi_b(t) \tag{3}$$

where ρ is the factor which controls the effect of the balancing process in the reward, μ_b is the responsibility factor for bucket b, $\omega(t)$ is the balancing ratio and $\xi_b(t)$ is the wholesale bidding performance.

Each bucket b has the responsibility factor μ_b which discounts its balancing cost responsibility:

$$\mu_b = \frac{1}{\gamma^{i_b - 1}} \tag{4}$$

where γ is the parameter which controls how the responsibility factor is distributed across buckets and the i_b is the ordinal number of a bucket b. As a consequence, this factor steers ABS towards minimizing the balancing cost for close proximities and optimizing wholesale bidding performance (i.e., buying cheaply and selling expensively) for far proximities.

The balancing ratio $\omega(t)$ is the normalized ratio between balancing energy[1] and distributed energy[2] given by:

$$\omega(t) = \begin{cases} \dfrac{balThreshold - balRatio(t)}{1 - balThreshold}, & \text{if } balRatio(t) \geq balThreshold \\[2ex] 1 - \dfrac{balRatio(t)}{balThreshold}, & \text{if } balRatio(t) < balThreshold \end{cases} \tag{5}$$

where $balRatio(t)$ is the ratio between balancing energy and distributed energy at time t:

$$balRatio(t) = \frac{|balancingEnergy(t)|}{distributionEnergy(t)} \tag{6}$$

[1] The balancing energy is defined as the final supply/demand imbalance in the observed timeslot.

[2] The distributed energy is defined as the total energy delivered for the observed timeslot. It is the sum of the positive net load of the broker's customers and the positive net export of energy through the wholesale market.

and $balThreshold$ is the maximum threshold for $balRatio(T)$. If the $balRatio(t)$ is higher than the maximum balancing threshold, the function will punish the action with negative rewards, otherwise the action is awarded with the positive reward.

The wholesale bidding performance ξ_b is introduced to prevent ABS from targeting the energy balance at all cost (i.e., engaging into destructive wholesale bidding):

$$\xi_b(t) = \frac{buyRatio_b(t) \times buyEnergy_b(t) + sellRatio_b(t) \times sellEnergy_b(t)}{buyEnergy_b(t) + sellEnergy_b(t)} \quad (7)$$

where $buyRatio_b(t)$ and $sellRatio_b(t)$ are the bucket's weighted mean buying price and weighted mean selling price for timeslot t, normalized over weighted mean wholesale price. The $buyEnergy_b(t)$ and $sellEnergy_b(t)$ are the bucket's total energy bought and sold for the timeslot t. The aforementioned mean prices are weighted based on the energy sold or bought. Since $buyEnergy_b(t)$ and $sellEnergy_b(t)$ may be sensitive to outliers (e.g., peaks in buying prices), we use the following sigmoid function

$$S(x) = \frac{2}{1 + \exp^{-x}} - 1 \quad (8)$$

to squash the values of $buyEnergy_b(t)$ and $sellEnergy_b(t)$ to acceptable [-1,1] range.

3 Adaptive Bidding System Evaluation

3.1 Experiment Setup

The preliminary results on the performance of the ABS are based on the data obtained from the Power TAC simulation environment. The evaluation process uses the following artefacts:

- Power TAC simulation platform for conducting experiments;
- ABS-enabled CrocodileAgent broker;
- CrocodileAgent brokers without ABS and named according to the strategy they use (e.g., *Generous* broker is the CrocodileAgent broker which only uses the *Generous* strategy);
- Extended version of the Power TAC Logfile Analysis Database[3] (PLA) for collecting both simulation data (e.g., market transactions and cash balances) and ABS-related data (e.g., bucket reportings) from experiments.

The experiment set includes scenarios where the ABS-enabled CrocodileAgent with the bucket sizes of 1, 8 and 24 compete in:

[3] The original version of the PLA database is developed by Markus Peters, Rotterdam School of Management, Erasmus University. The software is available on http://bitbucket.org/markuspeters/pla.

– Three players game against the broker with one of fixed strategies and the default broker;
– Five players game against all brokers with different fixed strategies and the default broker.

Table 2. Adaptive Bidding System configuration parameters used in experiments.

Parameter	Value
r	0.1
τ	1.0
e	0.2
$balancingThreshold$	0.15
ρ	2
γ	1.05

Each of scenarios was run twice and its duration was scheduled to last between 1000 and 1100 timeslots[4]. Regardless of the bucket size, the ABS-enabled agent used the same learning parameters listed in Table 2. It is important to note that those parameters were determined beforehand in a trail and that an optimality study of those parameters are beyond the scope of this paper.

3.2 Key Performance Indicators

Key Performance Indicators (KPIs) for balancing, wholesale and cash performances are used for the evaluation of the ABS. The exact summary values of KPIs are calculated against all data obtained from the set of experiments and they can be found later in Table 3 and Table 4.

Mean Balancing to Distribution Ratio The mean balancing to distribution ratio ($\mu_{BD}(b)$) is a metric which shows how did the agent b perform the balancing process in all experiments. This is an important KPI for the ABS since, according to the reward function $\pi_k(t)$, the primary objective of the ABS is to minimize the energy imbalance while keeping the cash flow on the wholesale market within reasonable limits (i.e., buys under and sells over the mean price). The lower the $\mu_{BD}(b)$, the better the balancing performance of the broker is.

Mean Cash The mean cash $\mu_{cash}(b)$ is the mean final cash balance agent b obtained per experiment. The greater the value, the more successful the agent is. It is important to notice that the majority of positive cash flow comes from the retail module which is out of the scope of this paper. However, this KPI provides a hint on how did wholesale activities influence the final outcome of the experiment.

[4] The duration of the experiment in Power TAC is not a fixed number. This prevents agents from scheduling unrealistic moves at the very end of the experiment.

Weighted Mean Buying and Weighted Selling Prices The weighted mean buying $\mu_{buyPrice}(b)$ and selling prices $\mu_{sellPrice}(b)$ provide good measurement on how much money did the agent b lose or gain per wholesale transaction. Those KPIs are used to determine whether the agent is paying too much for a good balancing ratio $\mu_{BD}(b)$ or why the agent is not able to keep the balancing cost under control (e.g., due to low buying orders). Mean values are weighted based on energy bought or sold, meaning that transactions with the low amount of traded energy have less influence on the resulting mean price and vice versa.

Mean Buying and Selling Energy The mean buying energy $\mu_{buyEnergy}(b)$ and selling energy $\mu_{sellEnergy}(b)$ tell how much energy did the agent b buy or sell per a wholesale transaction. Those KPIs provide additional clarification about agent's wholesale activities. For example, the low mean buying energy $\mu_{buyEnergy}(b)$ might prove that the agent's policy of using low mean buying price $\mu_{buyPrice}(b)$ did not pay off since it did not secure enough trades.

3.3 Results and Discussion

The previous subsection introduced six KPIs which were used for the evaluation of the ABS. This subsection offers a discussion of the results by examining each of the KPIs. Table 3 and Table 4 show calculated values of KPIs based on data from experiments. The values are considered for the following agents which participated in experiments:

- ABS_{24bs} - the ABS-enabled CrocodileAgent which has one bucket with the size of 24;
- ABS_{8bs} - the ABS-enabled CrocodileAgent which has three buckets with the size of 8;
- ABS_{1bs} - the ABS-enabled CrocodileAgent which has 24 buckets with the size of one;
- Default broker - the embedded broker with the simple behaviour on the retail market and wholesale market;
- Generous - the CrocodileAgent equipped with the *Generous* strategy;
- Neutral - the CrocodileAgent equipped with the *Neutral* strategy;
- Penny - the CrocodileAgent equipped with the *Penny* strategy;
- Stingy - the CrocodileAgent equipped with the *Stingy* strategy.

Results on μ_{BD} show that ABS-enabled agents, regardless of their bucket sizes, outperform agents which use *Penny* strategy and *Stingy* strategy in the balancing process. This is expected since both strategies try their best to buy the needed energy at the much lower price than the mean wholesale price. They will also attempt to sell the excess of energy at the much higher price than the price other agents are willing to buy for and thus many of their orders will be left unmatched. The *Default broker* was consistently losing the majority of customers and therefore it had only a limited impact on the rest of the market. The agent equipped with *Generous* strategy performed best during the balancing process.

Table 3. Key Performance Indicators for balancing and cash performance.

Agent	μ_{BD}	μ_{cash}[k€]
ABS_{24bs}	0.23	2,778
ABS_{8bs}	0.24	1,927
ABS_{1bs}	0.23	2,828
Default broker	0.37	9,6
Generous	0.22	1,920
Neutral	0.23	2,289
Penny	0.55	1,899
Stingy	0.71	1,292

Table 4. Key Performance Indicators for wholesale activities.

Agent	$\mu_{buyPrice}$ [€/MWh]	$\mu_{sellPrice}$ [€/MWh]	$\mu_{buyEnergy}$ $[MWh]$	$\mu_{sellEnergy}$ $[MWh]$
ABS_{24bs}	-42.15	26.53	5.05	-1.84
ABS_{8bs}	-41.25	25.93	4.00	-1.72
ABS_{1bs}	-42.79	26.80	5.51	-1.75
Default broker	-36.95	19.56	12.15	-10.81
Generous	-50.72	27.07	4.11	-1.46
Neutral	-27.24	42.88	3.84	-1.59
Penny	-23.63	37.76	1.93	-2.26
Stingy	-21.14	38.18	1.47	-1.89

This is also expected since the *Generous* strategy buys expensive energy and sells cheap energy. Therefore, its orders will be often matched and the agent will achieve the objective of keeping the balancing ratio low. The agent equipped with *Neutral* strategy is marginally better in the balancing process than the ABS-enabled agents. However, since the Neutral agent is constrained with only one strategy which closely follows the expected wholesale price, there is a serious doubt on the agent's performance in scenarios where the more diverse wholesale behaviour is required.

Although the balancing ratio is the most important aspect of the ABS, the good agent must also take care about its wholesale expenses. The weighted mean buying prices $\mu_{buyPrice}$ and weighted mean selling prices $\mu_{sellPrice}$, along with the mean energy bought $\mu_{buyEnergy}$ and mean energy sold $\mu_{sellEnergy}$ provide insights on wholesale activities. It is interesting to notice all ABS-enabled agents are having similar mean prices for buying and selling orders. Those prices are less favorable than prices offered by Neutral, Penny and Stingy agents, more favorable than selling prices offered by Default broker agent as well as more favorable than buying prices offered by the Generous agent. The biggest difference between ABS-enabled agents and other agents except Default broker is the fact that, according to $\mu_{buyEnergy}$, ABS-enabled agents are able to procure more energy then fixed agents. This in turn has an effect on the higher mean price of energy for ABS-enabled agents. The Penny agent, on average, has the most sales of energy due its buying policy (i.e., buying more energy than needed).

Mean cash μ_{cash} is the final metric used to show how the wholesale performance influence the overall agent's energy business. Since the retail market is the most lucrative place for making a profit in Power TAC, the μ_{cash} will largely depend on the agent's retail performance. The ABS-enabled agent with the bucket size of eight failed to beat the majority of agents. The proof of this are its values of μ_{cash} and $\mu_{buyEnergy}$ which are lower than other's ABS-enabled agents. All this may suggest the ABS configuration with three buckets is not the best way to setup the ABS-enabled agent. Still, it is encouraging fact that ABS-enabled agents with bucket sizes of 1 and 24 scored the best, meaning that adaptable bidding behaviour is better than fixed policy behaviour from competing agents. However, since the experiment setup has some limitations noted in the next subsection, all this observations should be taken with caution.

3.4 Experiment Limitations

Although experiment results show the effectiveness of the ABS, the authors are aware that the experiment setup has some limitations. First, the learning parameters of the ABS were chosen based on the results from other experiments not described in this work. Therefore, the experiment set should include the scenario with different learning parameters in order to present a rigorous analysis of their effect on the ABS performance. Second, the accuracy of inputs for the ABS is not measured. However, since ABS inputs (i.e., expected price and needed amount of energy) are calculated with simple prediction methods, they most likely contain noises which reflect the performance of ABS and thus the outcome of the experiment. Third, scenarios include the ABS-enabled CrocodileAgent and fixed-strategy brokers and therefore a more diverse scenarios with intelligent brokers are needed in order to thoroughly evaluate the proposed solution. All things considered, the evaluation part carried out in this work provide a preliminary analysis of the ABS.

4 Conclusion and Future Work

This paper presents an adaptive bidding mechanism (i.e., Adaptive Bidding System, ABS) for putting orders on the electricity wholesale market. We used the variant Roth-Erev reinforcement learning algorithm to coordinate wholesale bidding across different markets by choosing among four implemented wholesale strategies. Due to insufficient number of available agents in the official Power TAC broker repository, the experiment setup included modified versions of the CrocodileAgent. Specifically, all CrocodileAgent implementations used in the experiment had the same functionality of the retail module, while the wholesale module differentiated and included both single strategy versions as well as reinforcement learning enabled versions implemented through three different versions of the ABS. The experiment results show that the ABS-enabled CrocodileAgent outperforms the single strategy CrocodileAgent (i.e., without reinforcement learning) having in mind presented experiment limitations.

The future work will include: (i) determining optimal learning parameters, (ii) improving the accuracy of ABS inputs by using more advanced prediction methods, (iii) offering dynamic and more fine-grained strategies, and (iv) conducting similar experiments with more diverse brokers.

References

1. Babic, J., Matetic, S., Matijas, M., Buljevic, I., Pranjic, I., Mijic, M., Augustinovic, M.: The CrocodileAgent 2012: Research for Efficient Agent-based Electricity Trading Mechanisms. In: Special Session on Trading Agent Competition, KES-AMSTA 2012. Dubrovnik, Croatia (2012)
2. Babic, J., Podobnik, V.: An Analysis of Power TAC 2013 Trial. In: Trading Agent Design and Analysis, Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA. pp. 1–9 (2013)
3. Ketter, W., Collins, J., Reddy, P.: Power TAC: A competitive economic simulation of the smart grid. Energy Economics 39, 262–270 (Sep 2013), http://linkinghub.elsevier.com/retrieve/pii/S0140988313000959
4. Ketter, W., Collins, J., Reddy, P.P., Weerdt, M.D.: The 2013 Power Trading Agent Competition (May 2013), http://papers.ssrn.com/abstract=2268852
5. Peters, M., Ketter, W., Collins, J.: Autonomous Agents in Future Energy Markets: The 2012 Power Trading Agent Competition. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA. pp. 1298–1304 (2013), http://powertac.org/sites/default/files/6242-30892-1-PB-2.pdf
6. Peters, M., Ketter, W., Saar-Tsechansky, M., Collins, J.: A reinforcement learning approach to autonomous decision-making in smart electricity markets. Machine Learning pp. 5–39 (2013), http://download.springer.com/static/pdf/863/art\%3A10.1007\%2Fs10994-013-5340-0.pdf?auth66=1392229795_ad06fd4d8239e9b7291e102cf1f7a5ec\&ext=.pdf
7. Roth, A.E., Erev, I.: Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8(1), 164–212 (Jan 1995), http://www.sciencedirect.com/science/article/pii/S089982560580020X
8. Sun, J., Tesfatsion, L.: Dynamic Testing of Wholesale Power Market Designs: An Open-Source Agent-Based Framework. Computational Economics 30(3), 291–327 (Aug 2007), http://link.springer.com/10.1007/s10614-007-9095-1
9. Valogianni, K., Ketter, W., Collins, J.: Smart Charging of Electric Vehicles Using Reinforcement Learning. In: Trading Agent Design and Analysis, Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA. pp. 41–48 (2013)
10. Vytelingum, P., Cliff, D., Jennings, N.: Strategic bidding in continuous double auctions. Artificial Intelligence 172(14), 1700–1729 (Sep 2008), http://www.sciencedirect.com/science/article/pii/S0004370208000787