Electric Power Systems Research 124 (2015) 85–91

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Agent-based modeling of the demand-side system reserve provision

Edin Lakić[a,*], Gašper Artač[b], Andrej F. Gubina[c]

[a] BSP d.o.o., Dunajska 156, 1000 Ljubljana, Slovenia
[b] GEN-I, d.o.o., Krško, Slovenia
[c] University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 5 August 2014
Received in revised form 11 February 2015
Accepted 4 March 2015

Keywords:
Demand-side reserve provision
Agent-based modeling
Electricity market
SA-Q-learning
Economic costs
Benefits

ABSTRACT

Market simulators based on agent-based modeling techniques are frequently used for electricity market analyses. However, the majority of such analyses focus on the electricity markets bidding strategies on generation-side rather than on the demand-side. Meanwhile, the behavior of the demand-side in the system reserve provision has been less investigated. This paper presents a novel system reserve provision agent which is incorporated into a stochastic market optimization problem. The agent for the system reserve provision uses the SA-Q-learning algorithm to learn how much system reserve to offer at different times, while seeking to increase the ratio between their economic costs and benefits. The agent and its learning process are described in detail and are tested on the IEEE Reliability test system. It has been shown that incorporating the demand-side market strategies using the proposed agent improves the performance and the economic outcome for the consumers.

1. Introduction

Maintaining a high level of security of power system during operation continues to be a high priority in competitive electricity markets. The variable nature of renewable energy sources and demand combined with the possibility of a sudden loss of generating capacity require an adequate reserve in order to reduce the risk of load shedding. When these sources of variability compound they can lead to serious balancing problems, requiring an increase in the power reserve margin. The maintenance of power system security using only supply-side options will increasingly become technically more difficult and potentially more expensive [1].

A field of the demand response (DR) proposes a solution to this problem by introducing various demand response programs that could help enhance the demand-side participation at the electricity market [2,3]. In addition to a consumers' natural response to changes in electricity price, they may also be able to additionally modify their normal consumption in order to participate in the system reserve market [4]. Permitting and encouraging at least some retail consumers to face time-varying electricity prices provides economic, environmental, and reliability benefits. Action from DR technologies that are part of a portfolio of generators can be used to balance output and provide energy arbitrage opportunities, which could accrue as benefits to the portfolio owner or to intermittent generators using this resource [5]. Recent advances toward future intelligent grids (or smart grids, as they are known in Europe) enable the demand to respond quickly enough to provide some of the required system reserve. Day-ahead demand response program implemented as a source of spinning reserve can considerably reduce the total reserve cost and improve reliability indices [6]. The additional scheduling flexibility introduced by demand-side reserve offers brings significant gains in economic efficiency [7], and can lead to significant gains in social welfare in wholesale electricity market [8]. In recent literature regarding DR scheduling many different approaches were considered. Some of the authors adopted optimization techniques which usually minimize utility's cost to maximize social welfare [9–12]. Other authors adopted the ideas to random access protocols in data communications where the utility deals with customers' load demands which are sent to the utility and compete for the limited power generated over time [13,14]. Local demand response mechanisms can improve grid stability in island systems [15]. It can also be considered as possible business opportunity for the third party companies entering the domestic smart grid market which is affected by the changes due to advances in smart metering [16].

It is important to investigate whether a portion of the reserve requirements could be provided by demand, and to what extent. By providing reserve, consumers could maximize their benefits derived from electricity markets. To achieve this, consumers must determine their economic costs and benefits to decide if and when they should participate in the market. Their ability to provide reserve depends on their flexibility in modifying their

* Corresponding author. Tel.: +386 31 718 897; fax: +386 1 620 76 77.
E-mail address: lakicedo@gmail.com (E. Lakić).

http://dx.doi.org/10.1016/j.epsr.2015.03.003
0378-7796/

Nomenclature

Indices

b	index of season
D	demand
G	generation
j, k	index of loads and generating units
SRu	up-spinning reserve
SRd	down-spinning reserve
t	index of time (learning)
z	index of probability scenarios

Functions

$B_D(P_D)$	benefit function that loads derive from participating in the energy market
$c_R(r)$	cost function of employing reserve
$C_G(P_G)$	energy cost function
$C_R(R)$	reserve capacity cost function

Variables

a	action
EB^R	expected benefits that loads derive from participating in the reserve market
ESC	expected security cost participating in the reserve market
LC	involuntary load curtailment
P	scheduled real power
p_z	probability of scenario z
Q	Q-value
r	deployed reserve
R	scheduled reserve
$R^{\max}$	maximum reserve limit
S^{after}	net surplus after participating in the reserve market
S^{before}	net surplus before participating in the reserve market
$Temp$	the temperature in the Metropolis criterion
λ_R	market clearing price of security
σ	learning rate
ξ	random value
Γ	reward

Constants

n	number of states
ND, NG	number of consumers and number of generating units
NZ	number of probability scenarios
NT	number of scheduling hours
$P_{\max}, P_{\min}$	maximum and minimum forecasted consumption
$VOLL$	value of lost load
ζ	coefficient in Q-learning equations
φ	coefficient in the Metropolis criterion

consumption, the market prices of electricity and reserve services, and the frequency with which their reserve provision is called upon.

The demand-side strategies on the electricity market can be investigated using the agent-based modeling (ABM). This approach is able to capture a change in the behavior of different market participants and their adjustments to changes in the environment in which they operate. The application of the ABM combined with different reinforcement learning techniques has been rapidly developing in the area of power systems [17–19]. ABM is appropriate to apply to the systems where various entities i.e. agents, perform different actions but also take into consideration and modify their decisions based on previous experiences. This allows them to improve their performance to reach a certain goal.

The nature of the generation companies (GenCos) on one hand, and of the consumers (i.e. the demand-side) on the other, is appropriate for ABM approach because all players seek to adjust their behavior to market outcomes in order to improve their profits and benefits. However, most of the agent-based models were developed to investigate behavior of GenCos [20–24] although few also included analysis of the demand-side response [25–27].

To model and simulate the behavior of the GenCos Q-learning is typically used. In most cases it is used to evaluate how GenCos can raise their profits in the process of producing the electricity and offering it on the electricity market [21,24,28–31].

While various learning algorithms were applied to in different ABM electricity market simulators, a number of them argued for the application of the Q-learning [21,26,32]. In addition, several modifications of the Q-learning algorithm are being investigated, such as balancing between exploration and exploitation of an individual agent to reach the optimal policies for individual states, using the Metropolis criterion and the SA-Q-learning algorithm [29,30]. Also, almost all of electricity markets modeled diverse types of GenCos and used multi-agent system and learning approach [23]. They typically simulated either cooperative Q-learning [31], or take into consideration that each agent learned from its own individual learning experiences [28]. The results of ABM, together with reinforcement learning can be used for decision support systems in developing different strategies for submitting offers to the electricity market, as described in [33].

However, the majority of such analyses focus on the electricity markets bidding strategies, either on the generation-side or on the demand-side. To the knowledge of the authors of this paper, no attention has been paid to the ABM for the system reserve provision.

In the previous works we have presented new approaches to the stochastic modeling of the demand-side reserve provision in the co-optimised day-ahead electricity and reserve markets, where the consumer has the opportunity to participate as a reserve provider. Furthermore, the papers are focused on proposing a method to define the demand cost function for reserve provision. The method accounts for the costs and benefits that a consumer derives from its participation in the reserve market. The demand reserve offer function is determined by using optimization methods.

In contrast, in the proposed paper we present a novel agent for the demand-side system reserve provision. The intelligent agent is capable of learning how to improve its gain from its past actions. The agent learns how much system reserve to offer at different times, while seeking to increase the ratio between their economic costs and benefits.

The paper is divided in into the following sections: Section 2 describes the learning algorithm for the demand-side system reserve provision in detail, while in Section 3 the stochastic market model is explained. Section 4 presents the case study and the results are shown in Section 5. Finally, the conclusion is given in Section 6.

2. Learning algorithm for the demand-side system reserve provision

The objective of the demand agent is to maximize its reward, which in this case is defined as the difference between the expected benefit that consumer gains with the participation in system reserve provision and the expected security cost that consumers bear. The agent learns how to improve its gain by varying the amount of system reserve offered at different times, while seeking to increase the ratio between their economic costs and benefits. This is modeled using the SA-Q-learning approach, in which an

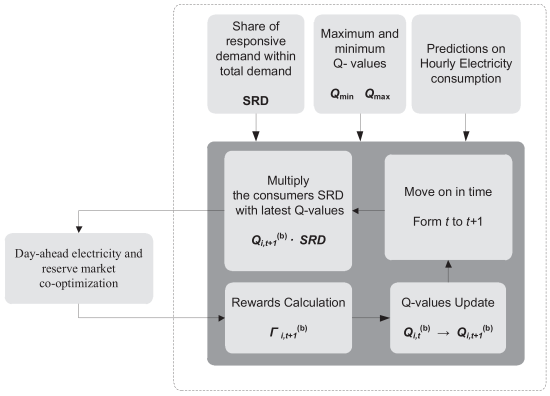

Fig. 1. Demand-side system reserve provision SA-Q learning flowchart.

agent can assume n different states, s_i ($i = 1, \ldots, n$) and can decide to carry out a single action, a_i ($i = 1, \ldots, n$) at each state. This enables the agent to move from state to state, with each state providing a reward to the agent. The following sub-section outline in more details the definitions of states, actions and rewards used for the demand-side system reserve provision agent modeling. They also describe the SA-Q-learning algorithm that enables the demand-side to improve its performance.

2.1. Algorithm flowchart

The demand-side system reserve provision SA-Q-learning procedure is illustrated in Fig. 1. Note that the rewards are calculated and the Q-values are updated for the entire next day (i.e. 24 h) within one cycle.

2.2. States

The states, s_i, describe all the circumstances in which the agent can find itself. In addition, states have to be chosen in such a way that moving from one state to another provides the agent with the right information regarding the benefits of such a move. Note that the state definition has to respond to the Markov property [34].

In the modeling of the demand agent presented here, the states are being defined according to the levels of the predicted consumption for the next operating period. These forecasted consumption define the interval $[P_{\min}{}^{(b)}, P_{\max}{}^{(b)}]$, where $P_{\min}{}^{(b)}$ and $P_{\max}{}^{(b)}$ are minimum and maximum forecasted consumption values for the considered time horizon. This interval is divided into the n segments, each covering such consumption width that agent visit each state equally frequent.

Furthermore, each of these n segments determines a single state s_i, and, considering the above discussion on different season, there would be n states for each of the seasons (or distinct time periods). Therefore, each state is denoted as $s_i{}^{(b)}$, while the action the agent can take when in this state are denoted as $a_i{}^{(b)}$, where ($i = 1, \ldots, n$) and ($b = 1, \ldots, 4$).

2.3. Actions

One individual action a_i for a single state s_i is defined. The SA-Q-learning equation provides Q-values for individual state s_i, and those are used for the share of responsive demand within total demand (SRD) shifting. In general, the SRD depends on the capability of the demand-side to increase or decrease its consumption within a specified period. The initial Q-value for all states at the beginning of the learning is 1. During the learning process these are being changed according to the reached rewards, i.e.,

$$a_i^{(b)} = Q_i^{(b)} \cdot SRD_i^{(b)} \tag{1}$$

where $Q_i{}^{(b)}$ is the factor for changing the demand level and the output of the SA-Q-learning process. If $Q_i{}^{(b)} > 1$, the $SRD_i{}^{(b)}$ will increase, and if $Q_i{}^{(b)} < 1$, the $SRD_i{}^{(p)}$ will decrease.

2.4. Rewards

The rewards are defined for each state in the following manner.

$$\Gamma_i^{(b)} = EB_{R\ i}^{(b)} \cdot ESC_{R\ t}^{(b)} \tag{2}$$

The benefits, EB_R, that consumer gains with participation in the system reserve provision in each hour are calculated as

$$EB_{D\ j,t}^{R} = \sum_{z=1}^{NZ} p_z \cdot \sum_{j=1}^{ND} ((S_{D\ j,t}^{after} - S_{D\ j,t}^{before}) + (income_{D\ j,t}^{SRu} + income_{D\ j,t}^{SRd})) \tag{3}$$

where

$$income_{D\ j,t}^{SRu} = \begin{cases} R_{D\ j,t}^{SRu} \cdot \lambda_{R\ t}, & \text{if } r_{D\ j,s,t}^{SRu} = 0 \\ r_{D\ j,s,t}^{SRu} \cdot c_{RD\ j,t}, & \text{if } r_{D\ j,s,t}^{SR\ u} \neq 0 \end{cases}$$
$$income_{D\ j,t}^{SRd} = \begin{cases} R_{D\ j,t}^{SRd} \cdot \lambda_{R\ t}, & \text{if } r_{D\ j,s,t}^{SRd} = 0 \\ r_{D\ j,s,t}^{SRd} \cdot c_{RD\ j,t}, & \text{if } r_{D\ j,s,t}^{SRd} \neq 0 \end{cases} \tag{4}$$

$S_D{}^b$ is the net load surplus before participating in the reserve market, $S_D{}^a$ is the net load surplus after participating in the reserve market, $R_D{}^{SRd}$ and $R_D{}^{SRu}$ are the scheduled demand down- and up-spinning reserve, λ_R is the market clearing price of security, $r_D{}^{SRd}$ and $r_D{}^{SRu}$ are the deployed demand down- and up-spinning reserve, and c_{RD} is the cost of employing demand down- and up-spinning reserve.

The expected security cost, ESC^R, that loads bear in each hour comprises the reserve cost and the involuntary load curtailment cost and is calculated as

$$ESC_{D\ j,t}^{R} = \sum_{z=1}^{NZ} p_z \cdot \left[\sum_{j=1}^{ND} (income_{D\ j,t}^{SRu} + income_{D\ j,t}^{SRd} + LC_{j,s,t} \cdot VOLL_{j,t}) + \sum_{k=1}^{NG} (income_{D\ k,t}^{SRu} + income_{D\ k,t}^{SRd}) \right] \tag{5}$$

where

$$income_{G\ k,t}^{SRu} = \begin{cases} R_{G\ k,t}^{SRu} \cdot \lambda_{R\ t}, & \text{if } r_{G\ k,s,t}^{SRu} = 0 \\ r_{G\ k,s,t}^{SRu} \cdot c_{RD\ j,t}, & \text{if } r_{G\ k,s,t}^{SR\ u} \neq 0 \end{cases}$$
$$income_{G\ k,t}^{SRd} = \begin{cases} R_{G\ k,t}^{SRd} \cdot \lambda_{R\ t}, & \text{if } r_{G\ k,s,t}^{SRd} = 0 \\ r_{G\ k,s,t}^{SRd} \cdot c_{RG\ k,t}, & \text{if } r_{G\ k,s,t}^{SRd} \neq 0 \end{cases} \tag{6}$$

$R_G{}^{SRd}$ and $R_G{}^{SRu}$ are the scheduled generation down- and up-spinning reserve, $r_G{}^{SRd}$ and $r_G{}^{SRu}$ are the deployed generation down- and up-spinning reserve, and c_{RG} is the cost of employing generation down- and up-spinning reserve, and LC is the expected involuntary load curtailment for particular consumer in each scenario. $income_{D\ j,t}^{SRu}$ and $income_{D\ j,t}^{SRd}$ is defined as per (4).

2.5. SA-Q-learning equation

In the paper, the agent is opportunistic, which means that it only considers current rewards, [28,34], and is modeled by setting the factor γ from the original Q-learning equation [35,36] to zero ($\gamma = 0$). Thus, Q-learning equation defines the new Q-value as the combination of the previous Q-value and the current reward, i.e.,

$$Q_{i,h+1}^{(b)} = \begin{cases} \left(1-\sigma_{i,h}^{(b)}\right)\cdot Q_{i,h}^{(b)} + \sigma_{i,h}^{(b)}\cdot \Gamma_{i,h}^{(b)}, & i = s_i^{(b)} \\ Q_{i,h}^{(b)}, & i \neq s_i^{(b)} \end{cases} \quad (7)$$

$$\sigma_{i,h}^{(b)} = \varsigma\cdot\sigma_{i,h-1}^{(b)};\quad \sigma_{i,h}^{(b)} = 1,\quad \varsigma < 1$$

where Q_{t+1} is the updated and Q_t is the previous Q-value, belonging to the state i; coefficient σ determines the weight that the previous reward has on the new Q-value – it equals 1 at the beginning of the learning process while shrinking with every visit of a particular state, according to the factor φ; R is the latest received reward [28].

Considering (7), it is important to observe that values of Q and Γ have to be of the same magnitudes, and thus have to be normalized before being used for calculation in the subsequent steps.

Unlike in the one-step Q-learning in the SA-Q-learning algorithm, when the agent selects actions, it does not only follow the policy learned so far, but also attempts to explore (according to the parameters such as temperature) by increasing chances of selecting actions other than those adopted by the current (possibly sub-) optimal policy [36]. In SA-Q-learning, the Metropolis criterion is used to make the solution escape from a local optimum, by changing corresponding probabilities for exploitation and exploration. The experiments in [36] demonstrate the advantages of the Metropolis criterion over other methods. The Metropolis criterion in SA-Q-learning can be described as follows:

1. Randomly select $a_{i,t}{}'^{(b)}$.
2. Select $a_{i,t}{}''^{(b)}$ according to the greedy policy where the strategy with taking the Q-value which yielded the highest reward so far is picked.
3. Generate a random value $\xi \in (0, 1)$

$$\xi < e^{[((Q_t^{(b)}(s_i^{(b)},a'^{(b)}_{i,t}) - Q_t^{(b)}(s_i^{(b)},a''^{(b)}_{i,t})))/Temp_{i,t}^{(p)}]} \quad a_{i,t}^{(b)} = a_{i,t}^{\prime(b)}$$

$$\text{otherwise} \quad a_{i,t}^{(b)} = a_{i,t}^{\prime\prime(b)} \quad (8)$$

Temp is the temperature as in the SA-Q algorithm. The geometric scaling factor criterion is used as the temperature dropping criterion in order to guarantee a slow decay of the temperature factor in the algorithm [36]. φ is usually constant which is close to 1.

$$Temp_{i,t}^{(b)} = \varphi_i\cdot Temp_{i,t-1}^{(b)} \quad t = 1, 2, \ldots \quad (9)$$

3. Market model

The used market model is formulated using a linear two-stage mixed-integer stochastic optimization program to formally incorporate the uncertainties into the model and to replicate a zonal day-ahead stochastic co-optimized energy and reserve markets [37–39]. The objective function is as follows:

$$\min \sum_{t=1}^{NT}\sum_{i=1}^{NG} C_{G\ i,t}(P_{G\ i,t}) - \sum_{j=1}^{ND} B_{D\ j,t}(P_{D\ j,t}) + \sum_{i=1}^{NG}(C_{RG\ i,t}^{SRu}(R_{G\ i,t}^{SRu}) + C_{RG\ i,t}^{SRd}(R_{G\ i,t}^{SRd})) + \sum_{j=1}^{ND}(C_{RD\ j,t}^{SRu}(R_{D\ j,t}^{SRu}) + C_{RD\ j,t}^{SRd}(R_{D\ j,t}^{SRd}))$$

$$+ \sum_{z=1}^{NZ} p_z \cdot \left[\sum_{i=1}^{NG}(c_{RG\ i,s,t}^{SRd}(r_{G\ i,s,t}^{SRu}) + c_{RG\ i,s,t}^{SRd}(r_{G\ i,s,t}^{SRd})) + \sum_{j=1}^{ND}(c_{RD\ j,s,t}^{SRu}(r_{D\ j,s,t}^{SRu}) + c_{RD,j,s,t}^{SRd}(r_{D,j,s,t}^{SRd}) + VOLL_{j,s,t}\cdot LC_{j,s,t})\right] \quad (10)$$

The first stage term is composed as follows. The first line treats power cost functions of generation $C_G(P_G)$ and demand-side $B_D(P_D)$. The second line contains the generation-offered up- and down-spinning reserve cost functions $C_{RG}{}^{SRu}(R_G{}^{SRu})$, $C_{RG}{}^{SRd}(R_G{}^{SRd})$. The third line contains the demand-offered up- and down-spinning reserve cost functions $C_{RD}{}^{SRu}(R_D{}^{SRu})$, $C_{RD}{}^{SRd}(R_D{}^{SRd})$. The remaining term defines second stage and shows the expected cost of providing security, containing the cost functions of deployed generator reserve $c_{RG}{}^{SRu}(r_G{}^{SRu})$, $c_{RG}{}^{SRd}(r_G{}^{SRd})$, the cost functions of deployed demand reserve $c_{RD}{}^{SRu}(r_D{}^{SRu})$, $c_{RD}{}^{SRd}(r_D{}^{SRd})$ and the cost of involuntary load curtailment, LC.

Some constraints relating to the first stage such as demand-supply power balance, generating unit and demand power technical constraints, contain no uncertainties. The remaining second stage constraints in the proposed model such as demand-supply power balance in each scenario, deployed spinning reserve limit and involuntary load curtailment limit are subject to uncertainties [37].

4. Case study

To demonstrate the SA-Q-learning algorithm, which is adopted to the system reserve provision agent, presented in Section 3 the IEEE Reliability test system [40] has been chosen, where the consumption is defined in Tables 2–4 in [40] for the period of one year, together with the duration and number of different seasons. The peak load is assumed to be 2850 MW. The IEEE Reliability test system has 32 generation units whose data are given in Tables 4–6 in [40]. The quadratic cost curve coefficients of generators are taken from [41].

To demonstrate the improvement in expected benefits and expected security cost for demand using agent based modeling of the demand-side system reserve provision, three different cases of SRD have been simulated. In Case 1, the SRD is set at 10%, in Case 2, it is set at 15%, and in Case 3, it is set at 20%. Each of these three cases was simulated with and without using the learning algorithm for the system reserve provision agent.

As discussed above, there is a limit on the level of the demand change. In this example, for the demand-side system reserve provision agent we consider that the SRD cannot increase above 50%, or decreased below 0%, in comparison to the originally hourly SRD values.

Initially, all Q-values $Q_{1,i}$ are set to 1, which means that there the agent does not have any prior knowledge that could influence its behavior.

5. Results

The results of the simulations are given for the three different cases, considering SRD set to 10%, 15% and 20%. The comparison between consumers with the demand-side learning is shown.

5.1. Demand-side reserve provision

Figs. 2–4 show the absolute values of the up-spinning reserve provided by the demand-side without (black line) and with the demand-side learning (gray line), for the period of one year and for

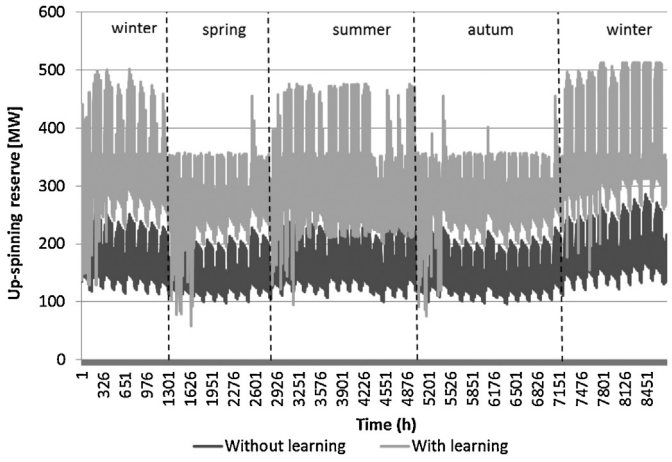

Fig. 2. Demand-side reserve provision for Case 1.

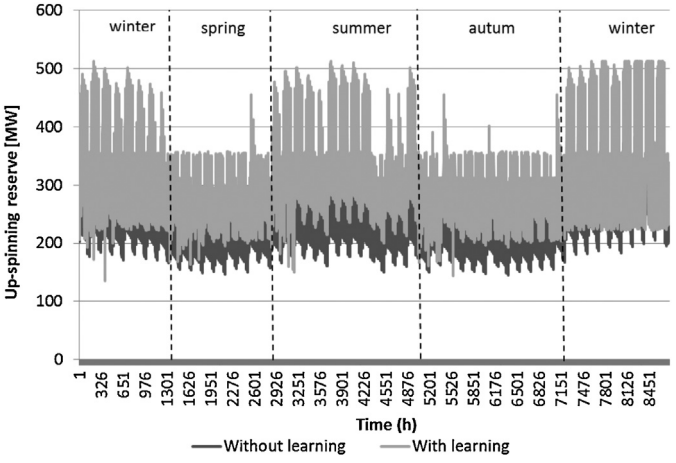

Fig. 3. Demand-side reserve provision for Case 2.

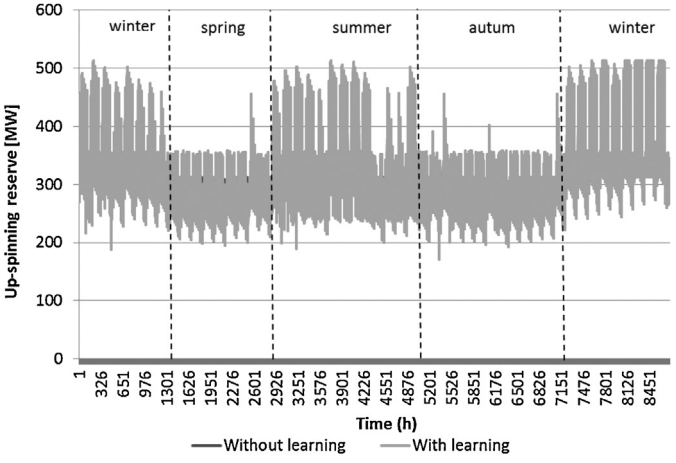

Fig. 4. Demand-side reserve provision for Case 3.

four different seasons. The year starts and finishes with the winter season.

In Fig. 2, when the SRD is set to 10%, without learning of the system reserve provision agent, SRD remains the same for the complete year, only the absolute values change according to the load profile. On the contrary, when the learning algorithm of the system reserve provision agent is applied, it is suggested to increase the given SRD. Note that the learning process re-starts at the beginning of each individual season. At those beginnings one can observe that the learning algorithm explores also the possibilities of lowering the SRD as to raise the consumers' benefit (negative gray spikes). This is due the applied SA-Q-learning algorithm, where in the begging of the learning process, the emphasis is more on the expiration to search within different possible *Q*-values for higher rewards. As time approaches to the end of the individual season, the SRD yet raises every time, which means that with the SRD set to 10% we are clearly under the optimal share for the demand-side reserve provision that the presented system can accept. Due to this fact, the system reserve provision agent suggests the increase of the SRD.

In Case 2, the SRD is set to 15% without learning of the system reserve provision agent, Fig. 3.

When observing the results without and with learning of the system reserve provision agent, a minor difference between Case 2 and Case 1 is obvious. Like in Case 1, in Case 2 reductions of the SRD at every beginning of each season can be seen due to the exploration of the learning process. At the end of the learning process it is suggested to raise the SRD to reach higher benefits, however not as significantly as in Case 1. This means, that with the SRD being set to 15% we are closer to the optimal share for the demand-side reserve provision for the given test-system as with the SRD set to 10%.

For this reason, in Case 3 we raise the SRD again to 20% without the learning of the system reserve provision agent, with the results presented in Fig. 4. While in Case 1 and in Case 2 the learning algorithm suggested the raise of the initial SRD in general, in Fig. 4, there is practically no change in the results with or without the learning of the system reserve provision agent. This means that with SRD set to 20% we are close to the marginal level of the SRD provision that the given system can accept. Further raising the SRD would neither bring any additional benefit to the consumers nor would the system be able to take so much reserves from the consumers.

By applying the learning to the system reserve provision agent we have come closer to the answer to the question what amount of the demand-side SRD is optimal for the given system.

5.2. Consumers reward

Figs. 5–7 show the difference between consumers' reward without and with the learning of the system reserve provision agent, for the period of one year. That means that consumers gain with learning algorithm if the difference is positive and lose if the difference is negative.

Fig. 5 shows the difference between consumers' reward without the learning algorithm and consumers reward with the learning algorithm of the system reserve provision agent for Case 1. One can

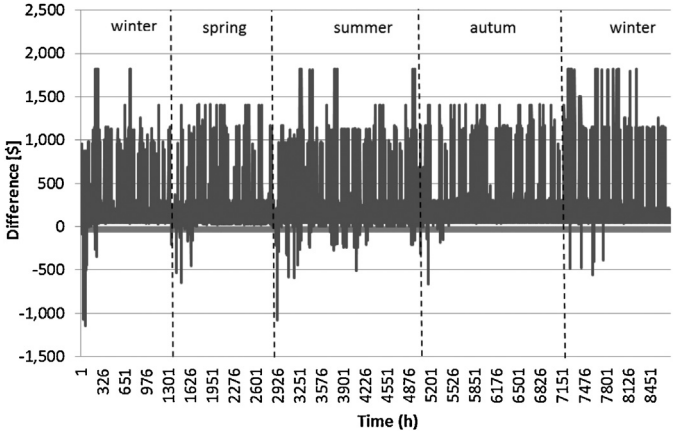

Fig. 5. Difference between consumers reward without the learning algorithm and consumers reward with the learning algorithm for Case 1.

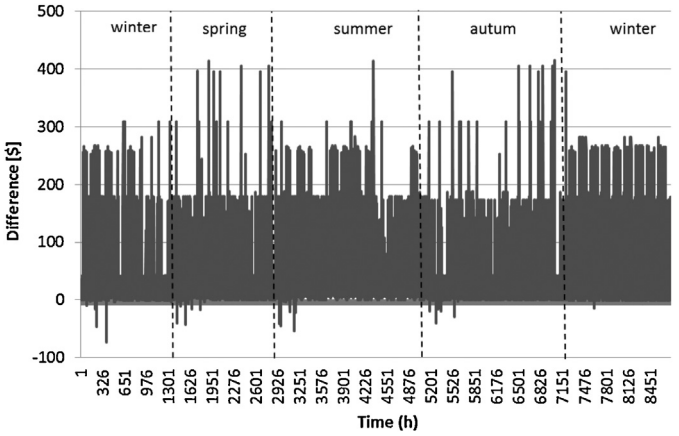

Fig. 6. Difference between consumers reward without the learning algorithm and consumers reward with the learning algorithm for Case 2.

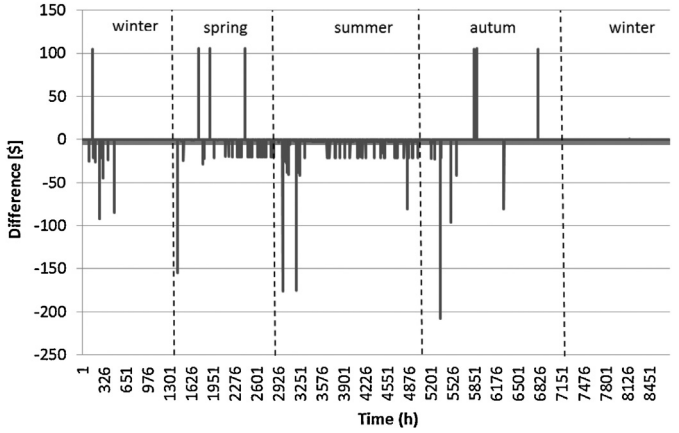

Fig. 7. Difference between consumers reward without the learning algorithm and consumers reward with the learning algorithm for Case 3.

observe that the reward rises significantly with the use of the learning algorithm. There is negative difference only at the beginning of the seasons. This is due to the exploration of the learning algorithm in the beginnings of the seasons, where the learning algorithm is exploring different actions.

Fig. 6 shows the difference between consumers' reward without the learning algorithm and consumers reward with the learning algorithm of the system reserve provision agent for Case 2. Similarly, there is also the reward rise with the use of the learning algorithm, although the raise of the reward is not as significant as it appears in Case 1.

Fig. 7 shows the difference between consumers reward without the learning algorithm and consumers reward with the learning algorithm for Case 3. One can observe that there are some negative spikes, especially in the beginnings of the seasons, where the learning algorithm is exploring different actions. Later on there are also some positive spikes when the reward rises with the use of the learning algorithm. Only in summer the difference is always negative, so the learning algorithm of the system reserve provision agent was not successful. However in Case 3 when the SRD is equal to 20% there is no such difference between the total reward without and with the learning of the system reserve provision agent.

In Table 1 the total consumers' reward with and without the learning algorithm of the system reserve provision agent in one year is shown. The reward is defined as the difference between the benefit and the security cost, and thus is negative. In Case 1 the reward raises significantly when applying the learning algorithm. In Case 2 it raises too, but not so extensively. In Case 3, the benefit

Table 1
Total reward of the consumers [€].

	Case 1	Case 2	Case 3
With demand-side learning	−6,741,410	−6,465,530	−6,421,769
Without demand-side learning	−9,128,220	−6,969,434	−6,418,729

is slightly lower when applying the demand-side agent, in comparison to the run without applying it. As mentioned above this is due to the exploration of the learning algorithm in the beginning of each seasons, when the learning algorithm is exploring with different actions, which are further away from the optimal solutions that the initial ones without learning. As has been stated in the 20% SRD is namely optimal from the consumers' point of view. With the SRD set to 20% we came very close to the optimal share for the demand-side reserve provision that the presented system can accept. Likewise, a similar result was presented in [30].

Furthermore, the rewards are growing with raising the SRD, be it without or with the demand side learning.

6. Conclusion

In the paper we present a novel agent-based approach that applies SA-Q-learning for the demand-side system reserve provision in co-optimized day-ahead electricity and reserve market. While Q-learning, as well as some other reinforcement learning techniques, was previously applied to simulate bidding strategies for GenCos or for demand-side, no attention has been paid to the demand-side system reserve provision and the share of responsive demand which can consumers devote for the system reserve provision.

The proposed system reserve provision agent builds on the capability of the SA-Q-learning technique to learn how to improve its gain by learning how much system reserve to offer at different times, while seeking to improve the rewarding ratio between the economic costs and benefits. In SA-Q-learning, the exploration of the learning algorithm in the beginnings of each season, where the learning algorithm is exploring with different actions is used to make the solution escape from a local optimum and to move closer to the global optimum. The detailed description of the system reserve provision agent uses to achieve the desired goals and its learning process are showed. This is illustrated on the IEEE Reliability test system, for which three different cases of SRD were simulated. The results indicate that the new agent improves performance and economic outcome for the consumers when it changes the SRD toward the optimal one. The extension of this research will investigate a possibility to upgrading the SA-Q-learning technique, as well as to apply it for the multi-agent system where different agents would represent different consumer groups at the electricity market. It means that each of the agents (i.e. consumers) may have its own strategy and limitations, and could also be able to get some learning experience from other agents.

Acknowledgments

This work was supported in within the framework of the Operational Programme for Human Resources Development for the period 2007–2013, 1. Development priorities Promoting entrepreneurship and adaptability, policy priorities 1.3: Scholarship Scheme. Its operation is partly financed by the European Union, European Social Fund.

References

[1] Y.T. Tan, D.S. Kirschen, Co-optimization of energy and reserve in electricity markets with demand-side participation in reserve services, in: IEEE PES Power Systems Conference and Exposition, 2007.

[2] M.H. Albadi, E.F. El-Saadany, A summary of demand response in electricity markets, Electric Power Syst. Res. 78 (11) (2008) 1989–1996.
[3] B. Kladnik, G. Artac, A. Gubina, An assessment of the effects of demand response in electricity markets, Int. Trans. Electr. Energy Syst. 23 (2013) 380–391, http://dx.doi.org/10.1002/etep.666.
[4] E. Hirst, Reliability benefits of price-responsive demand, IEEE Power Eng. Rev. 22 (November (11)) (2002) 16–21.
[5] G. Strbac, Demand side management: benefits and challenges, Energy Policy 36 (12) (2008) 4419–4426.
[6] E. Shayesteh, A. Yousefi, M. Parsa Moghaddam, A probabilistic risk-based approach for spinning reserve provision using day-ahead demand response program, Energy 35 (5) (2010) 1908–1915.
[7] J. Wang, N.E. Redondo, F.D. Galiana, Demand-side reserve offers in joint energy/reserve electricity markets, IEEE Trans. Power Syst. 18 (4) (2003) 1300–1306.
[8] M. Parvania, M. Fotuhi-Firuzabad, Demand response scheduling by stochastic SCUC, IEEE Trans. Smart Grid 1 (June (1)) (2010) 89–98.
[9] P. Faria, Z. Vale, J. Soares, J. Ferreira, Demand response management in power systems using particle swarm optimization, IEEE Intell. Syst. 28 (July (4)) (2013) 43–51.
[10] X. Zhang, G.G. Karady, S.T. Ariaratnam, Optimal allocation of CHP-based distributed generation on urban energy distribution networks, IEEE Trans, Sustain. Energy 5 (1) (2014) 246–253.
[11] K. Samarakoon, J. Ekanayake, N. Jenkins, Reporting available demand response, IEEE Trans. Smart Grid 4 (December (4)) (2013) 1842–1851.
[12] D. Li, J.K. Sudharman, Uncertainty Modeling and Price-based Demand Response Scheme Design in Smart Grid, 2014.
[13] S. Kishore, L. Snyder, Control mechanisms for residential electricity demand in smart grids, in: Proc. 1st IEEE Int. Conf. Smart Grid Comm, Gaithersburg, MD, USA, October, 2010, pp. 443–448.
[14] Y. Wang, I. Pordanjani, W. Xu, An event-driven demand response scheme for power system security enhancement, IEEE Trans. Smart Grid 2 (March (1)) (2011) 23–29.
[15] E. Kremers, J. Marı, O. Barambones, Emergent synchronisation properties of a refrigerator demand side management system, Appl. Energy 101 (2013) 709–717.
[16] S. Dave, M. Sooriyabandara, M. Yearworth, System behaviour modelling for demand response provision in a smart grid, Energy Policy 61 (2013) 172–181.
[17] S. McArthur, E. Davidson, Multi-agent systems for power engineering applications—part I: concepts, approaches, and technical challenges, IEEE Trans. Power Syst. 22 (4) (2007) 1743–1752.
[18] S. McArthur, E. Davidson, Multi-agent systems for power engineering applications—part II: technologies, standards, and tools for building multi-agent systems, IEEE Trans. Power Syst. 22 (November (4)) (2007) 1753–1759.
[19] I. Praca, C. Ramos, Z. Vale, M. Cordeiro, Mascem: a multiagent system that simulates competitive electricity markets, IEEE Intell. Syst. (2003) 54–60.
[20] G. Conzelmann, G. Boyd, V. Koritarov, Multi-agent power market simulation using EMCAS, in: Power & Energy Society General Meeting, 2005.
[21] N. Yu, C.-C. Liu, L. Tesfatsion, Modeling of suppliers' learning behaviors in an electricity market environment, in: International Conference on Intelligent Systems Applications to Power Systems, 2007.
[22] N.-P. Yu, S. Member, C.-C. Liu, J. Price, Evaluation of market rules using a multi-agent system method, IEEE Trans. Power Syst. 25 (1) (2010) 470–479.
[23] L. Gallego, O. Duarte, Strategic bidding in Colombian electricity market using a multi-agent learning approach, in: IEEE/PES Transmission and Distribution Conference and Exposition, 2008.
[24] A.C. Tellidou, A.G. Bakirtzis, S. Member, Agent-based analysis of capacity withholding and tacit collusion in electricity markets, IEEE Trans. Power Syst. 22 (4) (2007) 1735–1742.
[25] H. Oh, R.J. Thomas, Demand-side bidding agents: modeling and simulation, IEEE Trans. Power Syst. 23 (3) (2008) 1050–1056.
[26] K.Y. Lin, Dynamic pricing with real-time demand learning, Eur. J. Oper. Res. 174 (October (1)) (2006) 522–538.
[27] B. Kladnik, A. Gubina, G. Artac, K. Nagode, I. Kockar, Agent-based modeling of the demand-side flexibility, in: IEEE Power and Energy Society General Meeting, 2011.
[28] T. Hashiyama, S. Okuma, An electricity supplier bidding strategy through Q-learning, in: IEEE Power Engineering Society Summer Meeting, 2002.
[29] A. Rahimi-Kian, B. Sadeghi, R.J. Thomas, Q-learning based supplier-agents for electricity markets, in: IEEE Power Engineering Society General Meeting, 2005, pp. 2116–2123.
[30] J. Wang, Conjectural variation-based bidding strategies with Q-learning in electricity markets, in: 42nd Hawaii International Conference on System Sciences, 2009.
[31] M. Ahmadabadi, Expertness based cooperative Q-learning, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 32 (January (1)) (2002) 66–76.
[32] E. Beck, R. Cherkaoui, A. Germond, A comparison of Nash equilibria analysis and agent-based modelling for power markets, Int. J. Electr. Power Energy Syst. 28 (2006) 599–607.
[33] T. Sueyoshi, G. Tadiparthi, An agent-based decision support system for wholesale electricity market, Decis. Support Syst. 44 (2) (2008) 425–446.
[34] A. Weidlich, A critical survey of agent-based wholesale electricity market models, Energy Econ. 30 (4) (2008) 1728–1759.
[35] C.J.C.H. Watkins, P. Dayan, Technical note, Q-learning, Mach. Learn. 8 (1992) 279–292.
[36] M. Guo, Y. Liu, A new Q-learning algorithm based on the metropolis criterion, IEEE Trans. Syst. Man Cybern. Part B 34 (5) (2004) 2140–2143.
[37] G. Artac, D. Flynn, B. Kladnik, M. Hajdinjak, A.F. Gubina, The flexible demand influence on the joint energy and reserve markets, in: IEEE Power and Energy Society General Meeting, 2012, pp. 1–8.
[38] G. Artac, B. Kladnik, D. Dovzan, M. Pantos, A.F. Gubina, Demand-side system reserve provision in a stochastic market model, Energy Sources Part B: Econ. Plan. Policy (2015) (in press).
[39] G. Artac, D. Flynn, B. Kladnik, M. Pantos, A.F. Gubina, R. Golob, A new method for determining the demand reserve offer function, Electric Power Syst. Res. 100 (2013) 55–64.
[40] C. Grigg, et al., The IEEE reliability test system-1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee, IEEE Trans. Power Syst. 14 (3) (1999) 1010–1020.
[41] Q. Binh Dam, A.P. Sakis Meliopoulos, G.T. Heydt, A. Bose, A breaker-oriented, three-phase IEEE 24-substation test system, IEEE Trans. Power Syst. 25 (1) (2010) 59–67.