Decentralised demand response market model based on reinforcement learning

Miadreza Shafie-Khah[1], Saber Talari[2], Fei Wang[3], João P.S. Catalão[4] ✉

[1]*School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland*
[2]*Cologne Institute for Information Systems, University of Cologne, Cologne, Germany*
[3]*North China Electric Power University, Baoding 071003, People's Republic of China*
[4]*Faculty of Engineering of the University of Porto and INESC TEC, Porto, Portugal*
✉ *E-mail: catalao@fe.up.pt*

Abstract: A new decentralised demand response (DR) model relying on bi-directional communications is developed in this study. In this model, each user is considered as an agent that submits its bids according to the consumption urgency and a set of parameters defined by a reinforcement learning algorithm called Q-learning. The bids are sent to a local DR market, which is responsible for communicating all bids to the wholesale market and the system operator (SO), reporting to the customers after determining the local DR market clearing price. From local markets' viewpoint, the goal is to maximise social welfare. Four DR levels are considered to evaluate the effect of different DR portions in the cost of the electricity purchase. The outcomes are compared with the ones achieved from a centralised approach (aggregation-based model) as well as an uncontrolled method. Numerical studies prove that the proposed decentralised model remarkably drops the electricity cost compare to the uncontrolled method, being nearly as optimal as a centralised approach.

Nomenclature

Indices (sets)

$i(N)$ index (set) of hours
$t(T)$ index (set) of agents

Parameters

$D^t_{\max,DR}$	maximum potential of DR participation at hour t (kW)
$D^t_{\max,ac}$	maximum amount of loads to be supplied at hour t (kW)
$P^t_{\max,M}$	maximum amount of power to be purchased at hour t (kW)
D_C	total critical loads (kW)
D_L	total load demand (kW)
a_t	special action for Q-learning algorithm
α_t	learning rate for Q-learning algorithm
$P_{day,exp}$	expected amount of electricity to buy in a day (kW)
w	penalty factor
β_1	electricity price once no need to supply controllable loads (€/kWh)
β_2	sensitivity of the agents to pay for buying electricity
π^t_{ac}	price of selling power to customers from local DR market's viewpoint (€/kWh)
π^t_{MCP}	local DR market price (€/kWh)
π^t_{DR}	DR price from local DR market's viewpoint (€/kWh)

Variables

D_{TS}	amount of controllable loads (kW)
π_{state}	bidding for controllable loads (€/kWh)
L	load demand (kW)
D^t_{DR}	scheduled DR purchased from customers (kW)
D^t_{ac}	accepted amount of loads for agents to sell power (kW)
P^t_M	purchased power from local DR market (kW)
Q_t	Q-value at hour t for Q-learning algorithm
R_n	reward in Q-learning algorithm for agent n

$P_{day,real}$ electricity that is purchased in real by the agents during the day (kW)

1 Introduction

1.1 Motivation

Electricity consumption level is growing and rising, which caused some problems in electricity networks [1]. Hence, some measurements such as consumption reduction/shifting during peak period may be carried out to keep the energy balance with the lowest costs. Demand response (DR) programs enable end-users to modify the usual consumption and turn it into a cost-efficient pattern.

In the presence of electricity markets, customers are able to play an active role in a way that bids are based on their willingness to pay for electricity. Furthermore, with the existence of smart grid which include smart equipment such as advanced metering infrastructure and various communication facilities such as WiFi or Zigbee as well as Internet of Things potentials, customers are being able to accomplish two-way communications to utility for billing or monitoring [2]. Thus, all these facilities come up with the idea of considering customers as different DR agents who are able to bid actively in a competitive environment. Due to large number of end-users and to avoid computation burden, multi-agent systems (MASs) along with market-based control can be introduced from customers' side. Moreover, as decentralisation aims to make decisions based on the local needs, it helps to avoid irregular functionality of the market due to wrong decisions might be made by a central controller within wholesale market.

1.2 Literature review

Some articles have dealt with load participation in the electricity market. Customers in [3] are able to participate in electricity market through definition of a DR mechanism for improvement of the efficiency of renewable energy sources integration in the electric network. Papavasiliou *et al.* [4] schedule short-term energy consumption with considering an agent as an interface among end-users and the control market.

Some other studies have worked on various bidding strategies for customers. In [5], a downward temperature-price bidding function is presented for DR implementation, in a way that the electricity price is associated with the real temperature at home. A stepwise linearised profile via ten various fixed rates is introduced in [6]. An incentive bidding has been also suggested in [7, 8]. A linear bidding function has also proposed in [9].

A control signal of price is determined for ISO to motivate all players for modification of demand and supply in [10] by using a decentralised approach.

Based on [11], some DR programs (DRPs) are employed and designed for domestic end-users. However, expanding various types of technology with respect to smart grid, residential DR is being more applicable [12]. Implementation of DR in [13–16] is just practical through smart grid concept with smart facilities. Moreover, Shafie-Khah and Siano [17] describe the application of DRPs in a residence with aid of smart meters and advanced communication potentials.

According to [18], the smarter controlling systems of the domestic customers lead towards the more effective power management of end-users. Some other works [19] have studied the interaction among customers and upper utilities for the accomplishment of DR. DR can be applied in a pool-based market in [20]. Moreover, DR can be employed in the form of DR exchange (DRX) in day-head market [21, 22].

All the mentioned literatures have employed a central approach for cost-minimisation relying on a bi-directional communication between customer and upper aggregators that directly control end-users' DR potential. Besides privacy issues, for huge amount of customers, the central approach is not applicable due to the complexity and high computation burden. Therefore, decentralised control approach can be employed to overcome the problem [23].

To this end, customers' constraints can be considered in an aggregated way like [24–26] have accomplished. In [24], an agent approach has been applied for large number of appliances to participate in DR based on temperature constraints. A blockchain decentralised DR method has been used in [25] for financial settlement of DR based on consensus algorithm. Another multi-agent based approach has been performed in [26] for DR in microgrids in order to determine DR price and incentive through Lagrangian multiplier. However, once the influence of DR-enabled customers' bid is not considered, undesirable events such as avalanche effects, which is simultaneous reactions or errors can occur in foreseeing consumers' behaviour related to the price signals. To avoid these mentioned outcomes, bi-directional communications are considered in the new decentralised approach. Besides, applying bidding procedures [27], or iterative approaches [28], can overcome load synchronisation problems in case of rising the communication requirements. In [29], DR in virtual power plants can be traded within the DRX framework in a decentralised way in connection with intraday market. Likewise in [30], the interconnection of decentralised control of DR with bulk power market has been discussed.

There are not many studies that considers customers as several agents who are able to bid DR prices that can be optimised via a machine learning algorithm in a specific electricity market. The current study aims to cover and complete the previous studies in area. Outputs of this strategy can be useful for customers in different aspects such as minimisation of buying electricity.

1.3 Contributions

The intention of the current paper is to develop a decentralised scheme under consideration of responsive loads. In other words, the proposed model runs a market that DR-enabled end-users can bid based on their energy needs and the market responds them. Utilising a Q-learning algorithm, all bids can be optimised. End-users, as different agents, decide to purchase electricity cost-efficiency based on the local DR market clearing price (LDRMCP) as well as the previously determined demand bid curves. These curves have been considerably optimised via Q-learning algorithm in a convergent iterative process. All bids are sent to local markets that communicates the bids into the system operator (SO). Instead

of considering a DR aggregator, customers can buy electricity from the local market in a decentralised cost-efficient way by running a minimisation program for electricity purchase. In fact, in this paper, using an incentive-based DR program, customers can bid the DR quantity and price according to the optimised approach to define the best price. It means that customers bid based on the results of Q-learning algorithm.

The contributions of the paper are briefly as follows:

- Considering end-users as different agents who are able to bid actively in a decentralised market-based control scheme.
- Applying a Q-learning algorithm to optimise the customers' bids.
- Running an optimisation program by local DR market to maximise social welfare.

1.4 Organisation

The remainder of the paper is structured as below. Section 2 represents the bidding strategy including MAS, market scheme, bids aggregation and clearing method of local DR market applied for this approach. In Section 3, Q-learning algorithm and the implementation method in this work are introduced. Case study and numerical results are described in Section 4. The conclusion is in Section 5, and centralised method used for making a comparison among the proposed scheme and centralised one is presented in the Appendix.

2 Bidding strategy

2.1 Multi-agent system

MASs allow to manage the complex systems, which are modelled as groups of intelligent agents. MASs enables interacting among different agents and adapting their attitudes thoroughly. Hence, MASs would be used in a vast diversity of issues, ranging from market modelling to grid control and automation [31].

As mentioned, in the proposed model, end-users are considered as agents with their own local aims and are also able to bid in the market. Therefore, they need to interact with a market in order to work more efficiently. To this end, a market-based control scheme is presented as follows.

2.1.1 Market-based control scheme: The employed market-based scheme is on the basis of demand–supply models for price definition. Thoroughly, agents compete with each other through bidding to the market based on their willingness to purchase the electricity. Bids are sent to a local DR market where DR market price is cleared and after LDRMCP definition, it is communicated in the reverse direction to end-user. The market price serves as the control signal that is sent to customers to assign the resources. Thus, the amount of electricity to purchase is set based on equilibrium price and individual bids.

Fig. 1 presents the proposed scheme. The communication between the three considered levels of a MAS and the integration of players in the market-based control strategy are depicted.

2.1.2 Loads: Demand of end-users is divided into two categories including controllable and non-controllable (critical) loads. Non-controllable loads are not able to be managed and ought to be supplied all the time; otherwise, customers encounter major problems in their lifestyle such as refrigerator usage. While, controllable loads can be controlled by customer without facing any problem during the day, e.g. HVAC systems whose consumption can be scheduled within a suitable period of time.

A control function for each house is defined as follows [32]:

$$D_{\text{TS}} + D_{\text{C}} \leq D_{\text{L}} \tag{1}$$

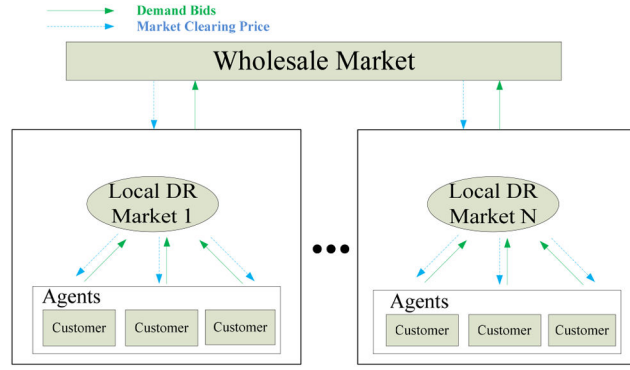$$D_{\text{TS}} = \sum_{i=1}^{N} D_{\text{TS}, i} \tag{2}$$

Fig. 1 *Communication between the MAS levels*

$$D_C = \sum_{i=1}^{N} D_{C,i} \qquad (3)$$

The variable D_{TS} represents the sum of controllable loads $D_{TS,\,i}$, D_C indicates the sum of non-controllable loads $D_{C,i}$ and D_L is the maximum power that the user can receive from the network. Thus, this formulation can model the willingness of each agent to buy electricity. It means each agent needs to purchase a volume of power within a bound, in which the minimum is critical loads and the maximum is total possible power to receive.

2.2 Demand bids

As mentioned earlier, unidirectional communication in the decentralised control schemes leads to obstacles for customers regarding DR bidding. Thus, for accomplishment of bidding process via each individual agent who is customer, bidirectional communication is applied.

The bid function is set in two intervals including inelastic and flexible intervals. The former is in respect with critical loads and the latter represents controllable loads. Two steps are considered in the second interval, which allows two degrees of freedom. The relevant stepwise linear curve is depicted in Fig. 2 and formulated in (4) as follows:

$$\text{demand (price)} = \begin{cases} \pi_{\max}, & 0 \le L \le D_C \\ \pi_{\text{state}}, & D_C \le L \le D_{\text{int}} \\ \pi_{\text{int}}, & D_{\text{int}} \le L \le D_L \end{cases} \qquad (4)$$

where

$$D_{\text{int}} = \frac{(D_L + D_C)}{2} \qquad (5)$$

Also, D_C and D_L represent the critical loads and maximum possible demands at each hour. π_{int} is introduced in (8). $\pi_{\max}$ indicates the maximum allowed price in the market that is allocated to critical demands. π_{state} is the DR price that end-users bid for controllable loads, which can be calculated as

$$\pi_{\text{state}} = \frac{t}{T} * \frac{D_C^t + D_{TS}^t}{D_L} \qquad (6)$$

where t denotes each hour in horizon time T and since the problem is in a day-ahead market T is 24 h. D_{TS}^t is the amount of controllable demands at each hour t in which the summation D_{TS} have to be supplied during the horizon time. In other words, once each agents buys electricity at each hour t, D_{TS} decreases because a part of D_{TS} has been covered through the purchased power. D_C denotes the volume of critical loads that an agent should satisfy. The intention of π_{state} is to define a DR price for agents who participate in DR based on their DR portion and possibility of load shifting within a day.
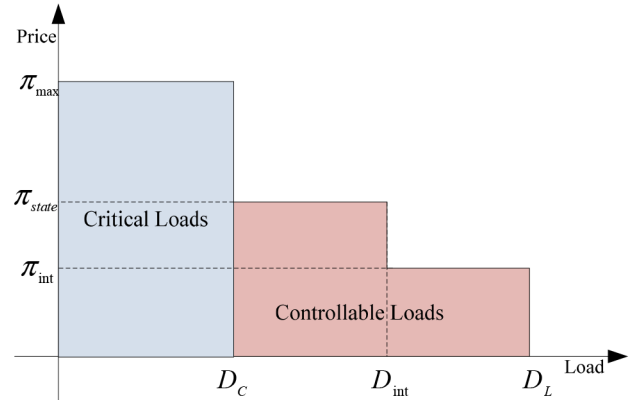

Fig. 2 *Demand bid curve*

Accordingly, if the agents do not purchase enough electricity to cover D_{TS} until the end of the day, they have to buy more electricity at the end of the day, which leads to higher π_{state}. The bidding curve is different for agents, since each agent has different willingness and attitude. For example, if two agents have the same decisions to supply their controllable loads at a special hour, π_{state} is higher for the agent that has higher critical loads.

Therefore, the bidding curve shown in Fig. 2 illustrates general patterns for agents, as each agent may have various statuses with different bidding profile. The bidding curve for agents who have no controllable loads is formed with only one block with maximum possible price why D_C is equal to D_L. On the other side, for agents without critical loads, bidding curve has two blocks due to the fact that D_C is zero.

In this method, π_{state} and D_{TS} are the variables to define the price, however sensitivity of the price and the electricity purchase through the market should be modelled in order to obtain optimum bidding.

2.2.1 Optimum demand bidding: To adapt the bidding curves of customers to the market price, some actions must be carried out to determine the sensitivity of electrical energy needs to the prices.

To obtain an optimum bidding in a pricing function, the cost of electricity purchase should depend on the level of consumption. Therefore, multiple pricing rates are assigned to customers based on their consumption level. The pricing model can consist of several parameters in which each one determines a level of pricing rate.

It is noteworthy that the number of parameters forms the bid blocks; however the extra number has a direct impact on convergence speed while optimising in learning algorithm, which will be discussed in Section 3.

Hence, just two parameters are considered in this work, which divide the controllable load part of bidding curve into two parts. Accordingly, the optimum amount of π_{state}, the second bid block, is calculated through (7) with obtaining the two parameters β_1 and β_2 via a learning algorithm. Moreover, the third bid block is calculated as (8)

$$\pi_{\text{state}} = \beta_1 + \beta_2 * \left(\frac{t}{T} * \frac{D_C^t + D_{TS}^t}{D_L} \right) \tag{7}$$

$$\pi_{\text{int}} = \frac{\pi_{\text{state}}}{2} \tag{8}$$

β_1 is the price that an agent considers low enough once the customer has no need to buy the electricity to supply controllable loads. In other words, this parameter is independent from the customers' consumption and can be considered as the lowest willingness of agents to pay for the electricity. On the other hand, β_2 can be interpreted as the sensitivity of the agents to pay for buying electricity in order to supply their controllable loads. It must be highlighted that the impact of the level of consumption on the price is modelled by this parameter.

A wide range of values can be assigned to these parameters. Modifying the parameters' volume leads to achieve the minimum operation cost or maximum social welfare. Therefore, with employing the learning machine algorithm, a wide range of numbers is automatically tested to find the best price, which leads to minimisation of the total cost of electricity purchase.

2.2.2 Local market: To avoid the huge computation burden and make the method practical, it is essential to introduce a local DR market that receives individual bids sending the integrated bids to the wholesale market. This local market is introduced to remove the role of aggregators in the decentralised approach. Aggregators are designed for centralised approach, while in a decentralised control algorithm a market scheme fits better to settle the DR trading. In fact, local DR market receives the DR bids from end-users and clears the market based on available bids. To this end, the demand bidding curves will be summed up horizontally as depicted in Fig. 3. In other words, demand loads and related price are defined by horizontal sum of various individual demand bidding curves. The final curve would have a stepwise linear function with the number of steps equals to the amount of various prices in every single demand bidding curve. It is noteworthy that each local DR market is introduced in the distribution level in this paper. The local market is in a close connection with wholesale market with exchanging the required data. For example, local market receives the MCP from wholesale market to use in its algorithm to obtain LDRMCP. Then, the LDRMCP will transfer to end-user (agents) to make the final decision about buying the electricity. Fig. 3 shows bidding curve in a local DR market, which is in charge of bids of three agents.

Thus, three agents send their bidding curves to this local DR market and they are finally formed like what has been depicted in the curve. Inflexible loads are summarised with one relative maximum price, while flexible loads are formed in a stepwise shape based on the number of agents in the local DR market.
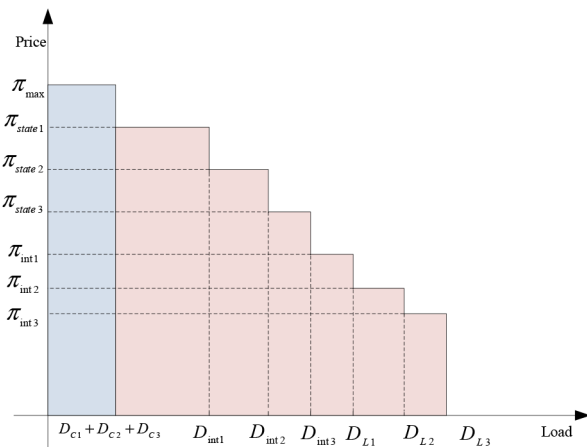

Fig. 3 *Bidding curve in local DR market for three agents*

2.2.3 Market clearing formulation: The proposed market strategy in this paper is based on pool market which gathers all supply and demand information to clear the market in a competitive way. In other words, all price signals control the responses to buy and sell DR and energy. It means consumers' bids have an effect on market prices, also after determination of LDRMCP, customers decide about power and DR.

Accordingly, an objective function is represented in (9) to maximise the social welfare from the local DR market viewpoint. Social welfare is the difference between customers' (agents') income and their cost. The first term in (9) indicates the accepted loads (D_{ac}^t) sold to all customers with the price (π_{ac}^t). The second term is the DR (D_{DR}^t) bought from all agents with the bid (π_{DR}^t). The third term is all power (P_M^t) bought from the local DR market with LDRMCP (π_{MCP}^t). Inequalities (10), (11) and (12) denote the limitation of demand, DR and power, respectively. Moreover, (13) is the balancing constraint for demand and supply. It is noted that decision variables are D_{ac}^t, D_{DR}^t and P_M^t

$$\text{Max} \sum_t \left(\sum_{ac} \pi_{ac}^t D_{ac}^t - \sum_{DR} \pi_{DR}^t D_{DR}^t - \sum_M \pi_{MCP}^t P_M^t \right) \tag{9}$$

subject to

$$0 \le D_{ac}^t \le D_{\max, ac}^t \tag{10}$$

$$-D_{\max, DR}^t \le D_{DR}^t \le D_{\max, DR}^t \tag{11}$$

$$0 \le P_M^t \le P_{\max, M}^t \tag{12}$$

$$\sum_{ac} D_{ac}^t - \sum_{DR} D_{DR}^t = \sum_M P_M^t \tag{13}$$

3 Reinforcement learning: Q-learning algorithm

There are three types of machine learning methodologies including: supervised, unsupervised and reinforcement learnings [33].

In supervised algorithms, labelled data are used to teach each agent, however in the unsupervised algorithms unlabelled data are utilised to teach agents. In the reinforcement algorithms, the learning process is to analyse the reward signal achieved by accomplishing a certain action [34]. Therefore, agents (customers in this paper) are able to find their optimum bidding strategy by this method with the interaction among electricity market.

The intention of the reinforcement learning is to maximise the rewards [35]. Thus, the algorithm tries to define the sequence of actions, which leads to obtain optimum rewards. According to this model, agents can conduct various actions which are defined as different prices for different load types [36].

There are several Q-values that are the expected rewards of possible actions on pair (β_1, β_2). In a particular action, a reward is assigned to this pair and preserved in a **Q**-matrix every time. Finding a strategy, which leads to maximisation of values in **Q**-matrix, called Q-value is the main goal of the proposed approach. Hence, Q-learning is employed to converge the action-value function Q to optimise the values. Since there are large number of customers and to decrease complexity, Q-values are considered independent from state of flexible loads consumption. Nevertheless, the state of flexible loads consumption are modelled easily in demand bid block in a way that π_{state} reflects the state of consumption for controllable loads. Thus, Q-function would be

$$Q_{t+1}(a_t) = Q_t(a_t) + \alpha_t(a_t) * [R_{t+1} - Q_t(a_t)] \tag{14}$$

where $Q_{t+1}(a_t)$ indicates the new Q-value or the updated Q-value by action a_t at the special hour. $Q_t(a_t)$ is the previous Q-value in that time. $\alpha(a_t)$ presents the learning rate, varying from 0 to 1 and determines the weight of new values compared with old values. This parameter has the key role in the convergence speed of this

approach. R_{t+1} is the reward obtained from implementing the action a.

All rewards are related to the action conducted by the agents in a particular iteration; therefore, they are independent from iteration. Accordingly, rewards are calculated as follows:

$$R_n = -\left[\sum_{t=1}^{24} \pi_{MCP}^t P_M^t + w\left(\left|P_{day,exp} - P_{day,real}\right|\right)\right] \quad (15)$$

In (15), the first term indicates the cost of purchasing electricity in a day. The second term presents the penalty for all agents that refuse to purchase sufficient or better-expected power in a day. This term helps to find the maximum of the reward faster through actions, because the actions caused to buy less electricity than expected are able to mislead the algorithm in a way that wrong actions can be pretended as higher relevant reward.

Likewise $P_{day,exp}$ indicates the expected amount of electricity to buy in a day, however $P_{day,real}$ represents the electricity that is purchased in real via the agents in the horizon time. In fact, $P_{day,exp}$ is a given data and is supposed to be available for each end-user based on historical demand data.

Considering penalty also aids the local DR market to compensate the small deviation between expected and real purchase power for agents. As once agents respond to the market based on their bids, they likely purchase more or less electricity than the expected one.

In the case of purchasing higher than expected, higher quantity is assigned to the local market price. Thus, agents' payments are based on the level of deviations. The **Q**-matrix can be formulated as follows:

$$Q = \begin{bmatrix} Q_{1,1} & \cdots & Q_{1,m} \\ \vdots & \ddots & \vdots \\ Q_{n,1} & \cdots & Q_{n,m} \end{bmatrix} \quad (16)$$

where the rows denote the agents and each column indicates a possible action. Moreover, n is the number of agents, while m would be the number of possible actions.

Q-learning algorithm needs a policy for making the decision regarding the best and most profitable action. Using ε-Greedy policy, agents enable to choose actions that maximise the reward each time. Accordingly, agents select a random available action a

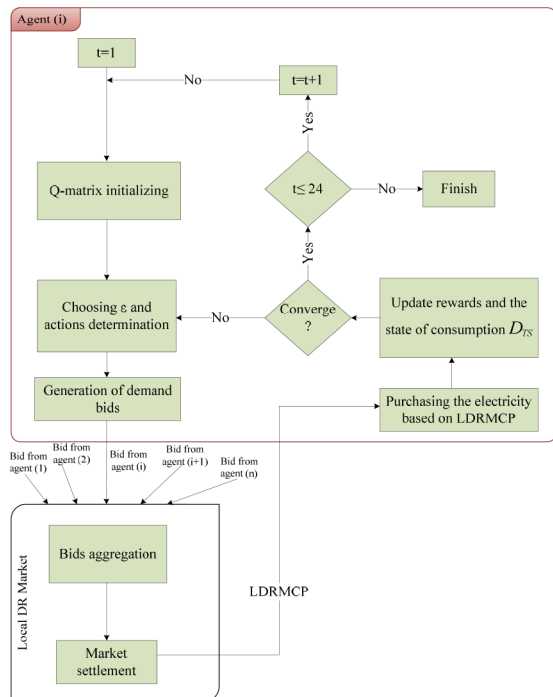out of all actions with the probability of $1 - \varepsilon$ and the action with the highest Q-value, or reward, is performed by agents with probability $1 - \varepsilon$. Therefore, the agents are able to check all actions and the relevant reward and then the highest related reward is selected.

Briefly, β_1 and β_2 are supposed to be optimised through assignment of different values based on reinforcement learning method known as Q-learning algorithm. Implementing this method, agents can realise their optimum bidding strategy with the interaction among the local DR market. The Q-values are the expected rewards for pairs (β_1, β_2) which are defined as the controllable loads price by updating in each iteration of the algorithm.

The different stages of implementing the proposed method in the paper is listed as follows and shown in Fig. 4. It is noted that convergence in this procedure occurs once the difference among two iterations is very low close to zero.

(i) *Performing actions*: From the first hour of day onward, the each action is performed via every single agent and determined based on ε-Greedy policy;
(ii) *Determining the bids*: The demand bids which represent the willingness of every single agent to purchase electricity would be determined;
(iii) *Sending bids to local DR market*: The bids of every single agent are sent to the local market;
(iv) *Clearing the electricity market*: All bids for flexible load and critical loads are collected with supply offer in the market to clear the electricity market at each hour.
(v) *Updating the consumption status*: Every single agent responds to price signals by updating the bids and consumptions. The reward related to the agent and the relative action are concurrently updated;
(vi) *Updating the **Q**-matrix*: **Q**-matrix is updated at each hour and new sets of actions are selected for all agents individually.

4 Case study and numerical results

4.1 Case studies

To evaluate the effectiveness of this approach, the outcomes are compared with those achieved from centralised approach as well as the outcomes achieved by the method disregarding DR. Both problems are solved in MATLAB and the computation time for centralised approach is 0.45 s and for decentralised one is 0.30 s. Moreover, to assess the impact of the DR participation portion on the outcomes, four DR participation portions are taken into account.

The first DR portion is 15% where DR launches having a remarkable effect on the market prices as well as the demand profile. The second and third participation levels equal to 30 and 60%, while in the fourth case study, DR is not considered; hence demand is taken inflexible into account.

Here, 100 customers are considered, and ε value in ε-Greedy algorithm is given 0.1 to encourage the algorithm for more exploration during the training period. Learning algorithm α is set to 0.65. The maximum bid price π_{max} is €3000 in this market. Meanwhile, the stepwise market bidding price for a day is in Fig. 5.
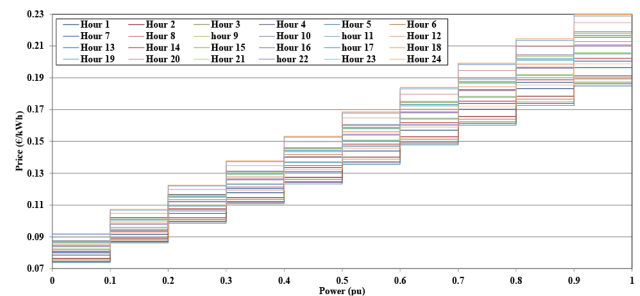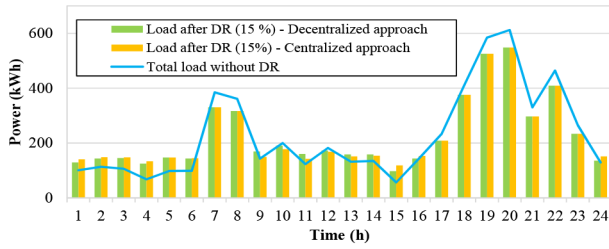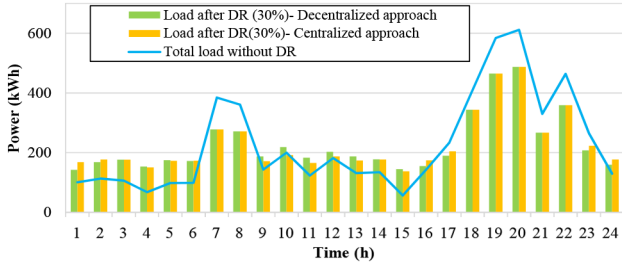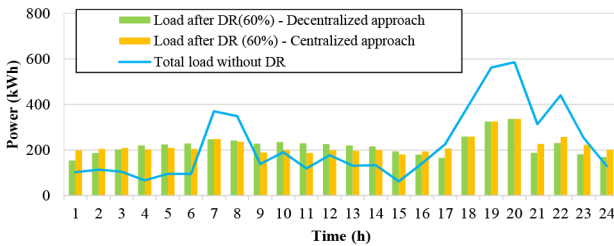

Fig. 4 *Flowchart of the proposed method implementation*

Fig. 5 *Market bidding price*

Fig. 6 *Results of total power purchased by customers in decentralised and centralised approaches in different cases*
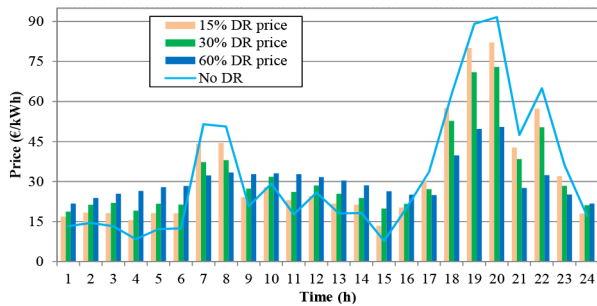*(a)* 15% DR participation, *(b)* 30% DR participation, *(c)* 60% DR participation

Fig. 7 *LDRMCP for different DR share percentage*

Table 1 Average cost of purchasing electricity

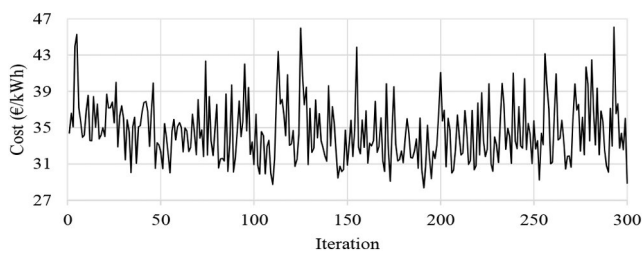| DR level	Cost – decentralised method, €/MWh	Cost – centralised method, €/MWh
0%	48.67	48.67
15%	27.47	25.74
30%	29.75	28.10
60%	31.57	29.58

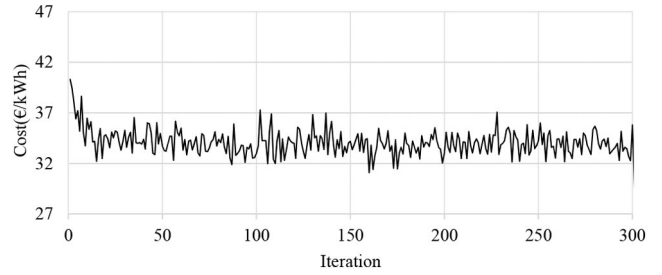**Fig. 8** *Iteration of CEA for 15% DR share*

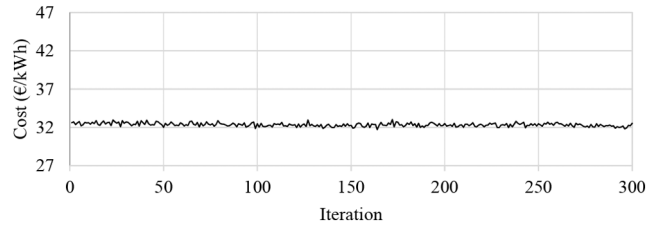**Fig. 9** *Iteration of CEA for 30% DR share*

Fig. 10 *Iteration of CEA for 60% DR share*

4.2 Consumption profiles

Fig. 6 shows customers load profile in different DR share levels and for uncontrollable loads. It represents the amount of electricity purchase in different DR states obtained from the proposed model. Results for 15% DR share in part (a), for 30% DR share in part (b) and 60% DR share in part (c) are depicted. As shown, the higher participation level of DR would be, the lower deviation among the minimum and maximum quantity of purchased electricity would happen.

Therefore, considering higher DR share caused more uniform electricity consumption during the day.

4.3 Market clearing price

Fig. 7 illustrates LDRMCP for all case studies including four DR portions. The LDRMCP is reduced during peak hours and would be higher during valley hours. LDRMCP in peak hours varies from around 90 to 50€/MWh in different cases. The reason is related to agents who participate in DRP buy power once the price is lower and there would be less competition for buying electricity in peak hours. Hence, increasing the number of DR participants, competition for buying electricity during valley hours rises followed by increasing price during such hours. However, higher portion of DR penetration leads to decrease in the variation of demand during the day, which causes a reduction in LDRMCP, considerably.

4.4 Average power cost

The average costs of electricity purchased by agents (CEA) in four case studies are compared in Table 1. This cost is calculated throughout the following formulation:

$$\text{CEA} = \frac{\sum_t \sum_n \pi_{ac}^{n,t} D_{ac}^{n,t}}{\sum_t \sum_n D_{ac}^{n,t}} \tag{17}$$

where n and t are the number of agents and hours. According to Table 1, it is concluded that as the customers' participation in DRP raises, the CEA increases as well. However, even for high portion of DR penetration, the CEA is considerably low compared with once there is no DR penetration. This result proves that the proposed model with DR can reduce CEA, remarkably. Moreover, the trend of CEA optimisation for 15, 30 and 60% DR penetration share in the iteration process are shown in Figs. 8–10. The cost would tend to change remarkably in iteration process once lower DR share is applied. For example, CEA for 15% DR share varies during the iteration process from 27 to 47€/MWh, while this element varies around 32 to 40€/MWh and 30 to 32€/MWh in 30 and 60% DR share, respectively.
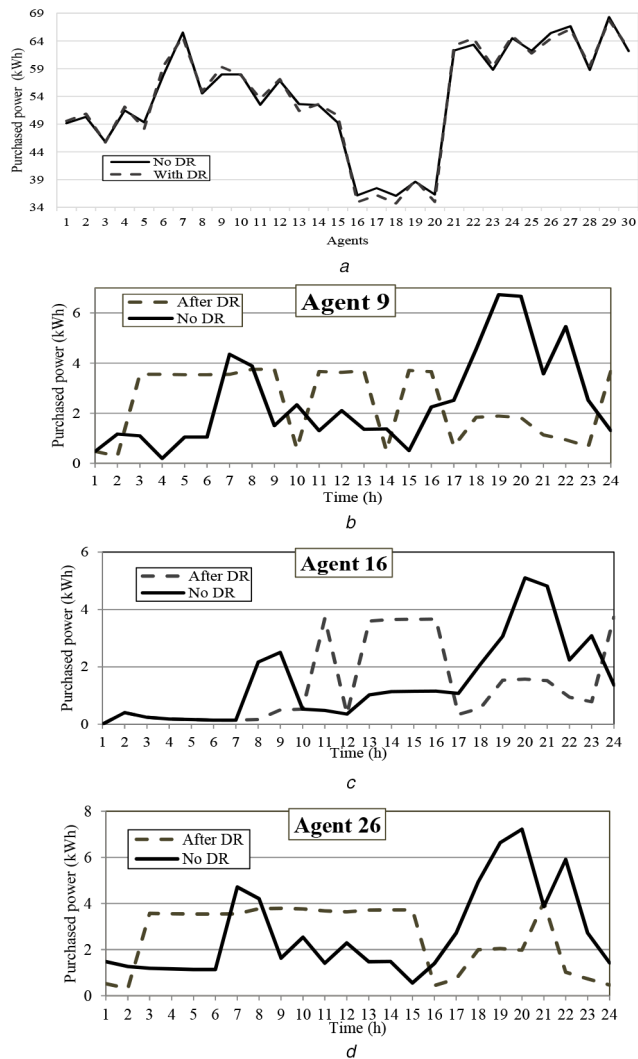
Fig. 11 *Comparison of purchased power for agents in 30% DR share with and without DR*

(a) Comparison of summation of purchased power in a day for 30 agents participated in DR, *(b)* Effect of DR on load profile of agent 9 during a day, *(c)* Effect of DR on load profile of agent 16 during a day, *(d)* Effect of DR on load profile of agent 26 during a day

The cost of purchasing electricity reaches about 30€/MWh in 15% DR penetration. For 30% DR share, the CEA without DR is 48.67€/MWh. However CEA, by using the proposed model, drops during the iterative process and reaches to 34€/MWh at the end of 300 iterations which can be seen in Fig. 9. The same attitude takes place for 60% DR portion in a way that the CEA would reach to 32.2€/MWh (Fig. 10).

4.5 Agents' behaviour

In Fig. 11, summation of consumption (purchased power) for 30 agents with considering DR and disregarding DR are compared. Moreover, load profiles of three samples of agents out of these 30 agents are depicted before and after decentralised DR implementation to illustrate the impact of the model on the consumption behaviour of each agent in detail. According to Fig. 11*a*, total consumption of an agent in a day can be less or more than once no DR is applied. For example, for agent number 16, the total consumption during a day after DR implementation is lower than before DR, although the consumption is optimised based on Fig. 11*c* in a way that most of peak-hour loads are shifted to the valley hours and others are totally curtailed. On the other hand, for the agent number 9 that the total consumption after running DR is more than before running DR, the load consumption in peak hours is reduced based on Fig. 11*b* and shifted to off-peak hours and also some new consumptions are scheduled for off-peak and valley
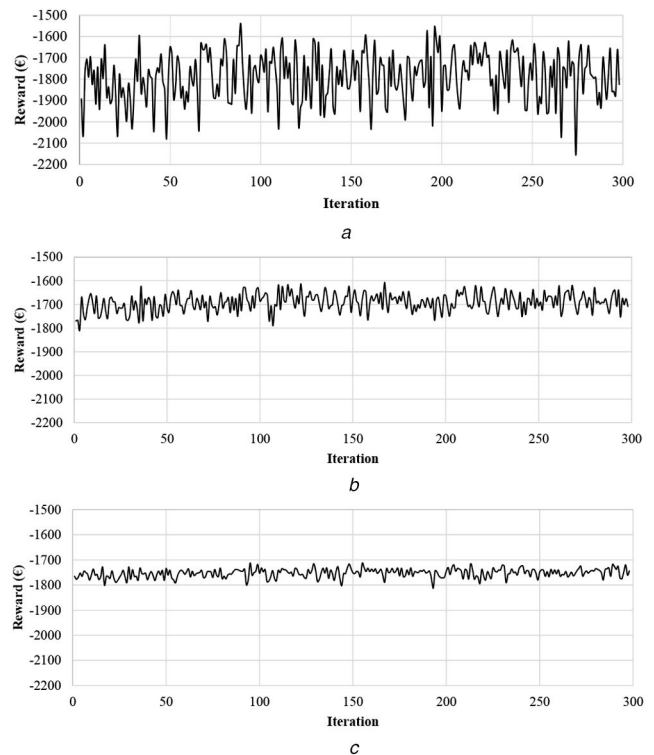
Fig. 12 *Variation of rewards during iteration process for three cases*
(a) 15% DR penetration, *(b)* 30% DR penetration, *(c)* 60% DR penetration

hours. This attitude for daily load profile is the same for the agent number 26 that no big difference is among total load consumption after and before DR application based on Fig. 11*d*.

4.6 Reward

In this section, the average variation of reward for all agents during the iteration process for three cases including 15% DR share, 30% DR share, and 60% DR share are demonstrated in Fig. 12. Based on Fig. 12*a–c*, the variation of reward in 15% DR penetration is higher than two other cases in a way that this volume varies among −€2200 to −€1500 in 300 iterations, while this variation is about €200 and €90 in 30 and 60% DR penetration, respectively.

4.7 Centralised versus decentralised model

In this part, the costs obtained throughout a centralised approach are compared with the ones obtained from the proposed model. The centralised model is presented in the Appendix. Indeed, impact of applying DR in centralised and decentralised model on total load profile is compared.

Moreover, CEA in centralised and decentralised models has a difference between 1.5 and 2€/MWh according to Table 1. Therefore, the results obtained by the employment of the proposed model are substantially similar to the ones achieved from the application of centralised method. It is verified that the results of the proposed model are approximately as optimal as the centralised model. Namely, local DR market has enough and completed information to bid to the market optimally on behalf of the agents in the proposed model.

5 Conclusions

A decentralised market-based scheme under consideration of DR was proposed. The results of the proposed framework have been compared with the case when there is no DR. Moreover, a comparison has been conducted among decentralised and centralised results. Proposing a bidding mechanism within a decentralised market-based control scheme is the main aim of the work. Accordingly, agents determine their optimum bids for buying electricity by employing the Q-learning algorithm. Therefore, electricity has been purchased based on the LDRMCP

and pre-defined demand bidding curve. To evaluate the efficiency of the method, four different percentages of DR penetration were considered. For all non-zero DR penetration perception, the model caused a decrease in the cost of electricity purchase compared with when no DR was considered. Moreover, agents who participate in DR buy more electricity during the hours with lower LDRMCP, because their consumption in peak hours has dropped. The proposed model not only caused a remarkable reduction on the electricity costs, but also decreased the deviation among the maximum and minimum amount of required electricity. Therefore, ascending the DR perception, this variation diminishes, which verifies the efficiency of the method in providing a more load balance in peak and off-peak hours. Comparing the results of the proposed method with the centralised method, it is transparent that both are approximately similar. For example, the differences of LDRMCP and load consumption between both methods are rather small. Thus, the proposed model is a reliable alternative to a centralised method because not only the results are very similar, but also it could provide easier and more scalable solutions for such complex problems.

6 Acknowledgment

7 References

[1] 'EIA – international energy outlook 2017'. Available at: https://www.eia.gov/outlooks/ieo/ (accessed Mar. 05, 2018)
[2] Talari, S., Shafie-Khah, M., Siano, P., *et al.*: 'A review of smart cities based on the internet of things concept', *Energies*, 2017, **10**, (4), p. 421, doi: 10.3390/en10040421
[3] Gungor, V.C., Sahin, D., Kocak, T., *et al.*: 'Smart grid technologies: communication technologies and standards', *IEEE Trans. Ind. Inf.*, 2011, **7**, (4), pp. 529–539, doi: 10.1109/TII.2011.2166794
[4] Papavasiliou, A., Hindi, H., Greene, D.: 'Market-based control mechanisms for electric power demand response'. 49th IEEE Conf. on Decision and Control (CDC), Atalanta, GA, USA, December 2010, pp. 1891–1898, doi: 10.1109/CDC.2010.5717572
[5] Siano, P., Sarno, D.: 'Assessing the benefits of residential demand response in a real time distribution energy market', *Appl. Energy*, 2016, **161**, pp. 533–551, doi: 10.1016/j.apenergy.2015.10.017
[6] Liu, Y., Holzer, J.T., Ferris, M.C.: 'Extending the bidding format to promote demand response', *Energy. Policy.*, 2015, **86**, pp. 82–92, doi: 10.1016/j.enpol.2015.06.030
[7] Sarker, M.R., Ortega-Vazquez, M.A., Kirschen, D.S.: 'Optimal coordination and scheduling of demand response via monetary incentives', *IEEE Trans. Smart Grid*, 2015, **6**, (3), pp. 1341–1352, doi: 10.1109/TSG.2014.2375067
[8] Hu, Q., Li, F., Fang, X., *et al.*: 'A framework of residential demand aggregation with financial incentives', *IEEE Trans. Smart Grid*, 2018, **9**, (1), pp. 497–505, doi: 10.1109/TSG.2016.2631083
[9] Li, N., Chen, L., Dahleh, M.A.: 'Demand response using linear supply function bidding', *IEEE Trans. Smart Grid*, 2015, **6**, (4), pp. 1827–1838, doi: 10.1109/TSG.2015.2410131
[10] Bahrami, S., Amini, M.H., Shafie-Khah, M., *et al.*: 'A decentralized electricity market scheme enabling demand response deployment', *IEEE Trans. Power Syst.*, 2017, **PP**, (99), pp. 1–1, doi: 10.1109/TPWRS.2017.2771279
[11] 'FERC: industries – reports on demand response & advanced metering'. Available at: https://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp (accessed Mar. 05, 2018)
[12] Li, Y., Ng, B.L., Trayer, M., *et al.*: 'Automated residential demand response: algorithmic implications of pricing models', *IEEE Trans. Smart Grid*, 2012, **3**, (4), pp. 1712–1721, doi: 10.1109/TSG.2012.2218262
[13] Talari, S., Shafie-Khah, M., Wang, F., *et al.*: 'Optimal scheduling of demand response in Pre-emptive markets based on stochastic bilevel programming method', *IEEE Trans. Ind. Electron.*, 2019, **66**, (2), pp. 1453–1464, doi: 10.1109/TIE.2017.2786288
[14] Parvania, M., Fotuhi-Firuzabad, M., Shahidehpour, M.: 'Optimal demand response aggregation in wholesale electricity markets', *IEEE Trans. Smart Grid*, 2013, **4**, (4), pp. 1957–1965, doi: 10.1109/TSG.2013.2257894
[15] Parvania, M., Fotuhi-Firuzabad, M., Shahidehpour, M.: 'ISO's optimal strategies for scheduling the hourly demand response in day-ahead markets', *IEEE Trans. Power Syst.*, 2014, **29**, (6), pp. 2636–2645, doi: 10.1109/TPWRS.2014.2316832
[16] Mahmoudi, N., Heydarian-Forushani, E., Shafie-Khah, M., *et al.*: 'A bottom-up approach for demand response aggregators' participation in electricity markets', *Electr. Power Syst. Res.*, 2017, **143**, pp. 121–129, doi: 10.1016/j.epsr.2016.08.038

[17] Shafie-Khah, M., Siano, P.: 'A stochastic home energy management system considering satisfaction cost and response fatigue', *IEEE Trans. Ind. Inf.*, 2018, **14**, (2), pp. 629–638, doi: 10.1109/TII.2017.2728803
[18] Ullah, I., Nadeem, J., Muhammad, I., *et al.*: 'A survey of home energy management for residential customers'. 2015 IEEE 29th Int. Conf. on Advanced Information Networking and Applications, Gwangiu, Republic of Korea, March 2015, pp. 666–673, doi: 10.1109/AINA.2015.251
[19] Carrasqueira, P., Alves, M.J., Antunes, C.H.: 'Bi-level particle swarm optimization and evolutionary algorithm approaches for residential demand response with different user profiles', *Inf. Sci.*, 2017, **418–419**, pp. 405–420, doi: 10.1016/j.ins.2017.08.019
[20] Nguyen, T., Negnevitsky, M., de Groot, M.: 'Pool-based demand response exchange: concept and modelling'. 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA, July 2011, pp. 1–1, doi: 10.1109/PES.2011.6039051
[21] Wu, H., Shahidehpour, M., Alabdulwahab, A., *et al.*: 'Demand response exchange in the stochastic day-ahead scheduling with Variable renewable generation', *IEEE Trans. Sustain. Energy*, 2015, **6**, (2), pp. 516–525, doi: 10.1109/TSTE.2015.2390639
[22] Saebi, J., Javidi, M.H., Oloomi Buygi, M.: 'Toward mitigating wind-uncertainty costs in power system operation: a demand response exchange market framework', *Electr. Power Syst. Res.*, 2015, **119**, pp. 157–167, doi: 10.1016/j.epsr.2014.09.017
[23] Najafi, S., Talari, S., Gazafroudi, A.S., *et al.*: 'Decentralized control of DR using a multi-agent method'. Sustainable Interdependent Networks, Springer, Cham, 2018, pp. 233–249
[24] Duong, N.H.S., Maillé, P., Ram, A.K., *et al.*: 'Decentralized demand response for temperature-constrained appliances', *IEEE Trans. Smart Grid*, 2019, **10**, (2), pp. 1826–1833, doi: 10.1109/TSG.2017.2778225
[25] Pop, C., Cioara, T., Antal, M., *et al.*: 'Blockchain based decentralized management of demand response programs in smart energy grids', *Sensors*, 2018, **18**, (1), p. 162, doi: 10.3390/s18010162
[26] Sakurama, K., Miura, M.: 'Communication-based decentralized demand response for smart microgrids', *IEEE Trans. Ind. Electron.*, 2017, **64**, (6), pp. 5192–5202, doi: 10.1109/TIE.2016.2631133
[27] Kok, J.K., Scheepers, M.J.J., Kamphuis, I.G.: 'Intelligence in electricity networks for embedding renewables and distributed generation'. Intelligent Infrastructures, Springer, Dordrecht, 2010, pp. 179–209
[28] Paverd, A., Martin, A., Brown, I.: 'Privacy-enhanced bi-directional communication in the smart grid using trusted computing'. 2014 IEEE Int. Conf. on Smart Grid Communications (SmartGridComm), Venice, Italy, November 2014, pp. 872–877, doi: 10.1109/SmartGridComm.2014.7007758
[29] Nguyen, H.T., Le, L.B., Wang, Z.: 'A bidding strategy for virtual power plants with the intraday demand response exchange market using the stochastic programming', *IEEE Trans. Ind. Appl.*, 2018, **54**, (4), pp. 3044–3055, doi: 10.1109/TIA.2018.2828379
[30] Durvasulu, V., Hansen, T.M.: 'Benefits of a demand response exchange participating in existing bulk-power markets', *Energies*, 2018, **11**, (12), p. 3361, doi: 10.3390/en11123361
[31] McArthur, S.D.J., Davidson, E.M., Catterson, V.M., *et al.*: 'Multi-agent systems for power engineering applications #x2014;part I: concepts, approaches, and technical challenges', *IEEE Trans. Power Syst.*, 2007, **22**, (4), pp. 1743–1752, doi: 10.1109/TPWRS.2007.908471
[32] Shao, S., Pipattanasomporn, M., Rahman, S.: 'Demand response as a load shaping tool in an intelligent grid with electric vehicles', *IEEE Trans. Smart Grid*, 2011, **2**, (4), pp. 624–631, doi: 10.1109/TSG.2011.2164583
[33] Murphy, K.P.: '*Machine learning: a probabilistic perspective*' (MIT Press, Cambridge, MA, 2012)
[34] Sutton, R.S., Barto, A.G.: '*Learning: an Introduction*' (MIT Press, Cambridge, MA, 1998)
[35] Krause, T., Andersson, G.: 'Evaluating congestion management schemes in liberalized electricity markets using an agent-based simulator'. 2006 IEEE Power Engineering Society General Meeting, Montreal, Quebec, Canada, 2006, p. 8, doi: 10.1109/PES.2006.1709123
[36] Rahimiyan, M., Mashhadi, H.R.: 'An adaptive Q-learning algorithm developed for agent-based computational modeling of electricity market', *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, 2010, **40**, (5), pp. 547–556, doi: 10.1109/TSMCC.2010.2044174

8 Appendix

8.1 Centralised model

In order to assess the proposed decentralised approach in this work, the results are compared with the results of a centralised approach.

In the proposed framework, individual agents directly participate in the local DR market by submitting their bids and respond to the market clearing price according to them. While in a centralised model, bidding is managed by a DR aggregator. In other words, DR aggregators bid into the market directly on behalf of customers. Therefore, in the centralised approach, DR aggregator plays the main role to implement the DR and is an interface among customers and the market.

The contract among DR aggregators and customers is in a way that DR aggregators bid to the customers based on the evaluation accomplished on customers' capabilities by transferred data from

customers to DR aggregators. Then, aggregators run DR contracts in the market to determine optimal DR offer by maximisation of their profit, and this data will be sent to SO.

Accordingly, the centralised objective function is presented as (18) which maximises DR aggregator's profit. DR aggregator income is to sell DR to market with considering MCP as the selling price and the cost of DR aggregator can be DR purchase from customers with predefined π_{DR}^t

$$\text{Max} \sum_t \pi_{MCP}^t D_{DR}^t - \pi_{DR}^t D_{DR}^t \qquad (18)$$

subject to

$$0 \le D_{DR}^t \le D_{max,DR}^t \qquad (19)$$

where (19) is limitation of DR assignment.