This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Agent-based Modeling and Simulation of the Electricity Market With Residential Demand Response

Shuyang Xu, *Student Member, IEEE*, Xingying Chen, *Senior Member, IEEE*, Jun Xie, *Member, IEEE*, Saifur Rahman, *Fellow, IEEE*, Jixiang Wang, *Student Member, IEEE*, Hongxun Hui, Tao Chen

***Abstract*—Nowadays critical peak load caused by residential customers has attracted utility companies and policymakers to pay more attention to residential demand response (RDR) programs. In typical RDR programs, residential customers react to the price or incentive-based signals, but the actions may naturally fall behind the flexible market situations. For those residential customers equipped with smart meters, they may contribute more DR loads if they can participate in DR events in a proactive way. In this paper, we propose a comprehensive market framework in which residential customers can provide proactive RDR actions in a day-ahead market (DAM). We model and evaluate the interactions between generation companies (GenCos), retailers, residential customers, and the independent system operator (ISO) via an agent-based modeling and simulation (ABMS) approach. The simulation framework contains two main procedures—the bottom-up modeling procedure and the reinforcement learning (RL) procedure. The bottom-up modeling procedure models the residential load profiles separately by household types to capture the RDR potential difference in advance so that residential customers may provide automatic DR actions rationally. Retailers and GenCos optimize their bidding strategies via the RL procedure. The modified optimization approach in this procedure can prevent the training results from falling into local optimum solutions. The ISO clears the DAM to maximize social welfare via Karush-Kuhn-Tucker (KKT) conditions. Based on realistic residential data in China, the proposed models and methods are verified and compared in a large multi-scenario test case with 30,000 residential households. Results show that proactive RDR programs and interactions among the market entities may yield significant benefits for both the supply and demand sides. The models and methods in this paper may be used by utility companies, electricity retailers, market operators, and policymakers to evaluate the consequences of the proactive RDR and the interactions among multi-entities.**

***Index Terms*—Residential demand response (RDR), agent-based modeling and simulation (ABMS), reinforcement learning (RL), electricity market.**

NOMENCLATURE

Abbreviations:

IoT	Internet of things.
DR	Demand response.
IDR	Industrial demand response.
CDR	Commercial demand response.
RDR	Residential demand response.
CPP	Critical peak pricing.
TCL	Thermostatically controlled load.
GenCo	Generation company.
ISO	Independent system operator.
CA	Consumer agent.
RA	Retailer agent.
GA	GenCo agent.
DAM	Day-ahead market.
ABMS	Agent-based modeling and simulation.
HEMS	Home energy management system.
RL	Reinforcement learning.
KKT	Karush-Kuhn-Tucker conditions.

Sets:

N	Set of household types, indexed by n.
K	Set of appliances, indexed by k.
V	Set of air conditioners, indexed by v.
H	Set of households, indexed by h.
T	Set of time interals, indexed by t.
R	Set of RAs, indexed by r.
G	Set of GAs, indexed by g.
D	Set of outloops, indexed by d.
E	Set of innerloops, indexed by e.
J	Set of agents, indexed by j.
A_j	Set of action space for agent j, indexed by i.

Variables:

$\Gamma_{n,h,k}(t)$	Operation time of the appliance k, household h, household type n during time interval t.
$T_{n,h,\text{set}}(t)$	Cooling set temperature of household h, household type n during time interval t.
$O_{n,h,v}(t)$	ON/OFF variable of air conditioner v, household h, household type n during time interval t.
w_r	Electricity sold by RA r.
μ_r	Markdowen rate for RA r.
q_g	Electricity supplied by GA g.
m_g	Markup rate for GA g.
$\lambda_1,\lambda_2,\beta_{1g},\beta_{2g},\beta_{3r},\beta_{4r}$	KKT multipliers.
π	Market clearing price.
$p_{j,d}(i)$	Selection probability of action i for agent j in training loop d.
$F_{j,d}$	Policy space for agent j in training loop d.
$q^*_{j,d,e}$	Simulated result of electricity supplied by GA j in training loop d,e.
$w^*_{j,d,e}$	Simulated result of electricity sold by RA j in training loop d,e.
$\pi^*_{d,e}$	Simulated market clearing price in training loop d,e.

DOI: 10.17775/CSEEJPES.2019.01750

Parameters:

$P_{n,h,k}$	Fixed power of the appliance k, household h, household type n.
$\gamma_{n,h,\mathrm{v}}$	Air conditioner refrigeration efficiency parameter of air conditioner v, household h, household type n.
p_1^n, p_2^n, p_5^n p_3^n, p_4^n, p_6^n	Price thresholds for lighting, plug, and air conditioner loads of household type n to take DR actions.
α, β, σ	Load reduction limits for lighting, plug, and air conditioner loads.
$\varphi_r, \phi_r, \omega_r$	Bid utility function parameters for RA r.
w_r^{max}, w_r^{min}	Upper/lower selling limit parameters for RA r.
a_g, b_g, c_g	Bid cost function parameters for GA g
q_g^{max}, q_g^{min}	Upper/lower supply limit parameters for GA g.
ε	A small parameter to test the price difference between two simulated results.
θ_j	Amount of the bidding actions for agent j.
ρ	Learning intensity parameter for agent j.

Functions:

$L_{n,h,k}$	Energy consumption of the appliance k, household h, household type n.
$L_{n,h,v}$	Energy consumption of the air conditioner v, household h, household type n.
$L_{n,h}$	Energy consumption of household h, household type n.
L	Energy consumption of household groups.
$\check{S}_r$	Strategic bidding utility function for RA r.
S_r	Real utility price function for RA r.
$\check{C}_g$	Strategic bidding cost function for GA g.
$B_{j,d,e}$	Profit for agent j in training loop d, e.
$\bar{B}_{j,d}$	Average profit for agent j in training loop d.
$R_{j,d}$	Reward for agent j in training loop d.

I. Introduction

A. Background and Research Motivations

Demand response (DR) programs have been proven to offer a variety of financial and operational benefits for both the supply and demand sides [1]-[2]. Industrial and commercial customers take the majority of DR providers, but the rapid growth of residential electricity consumption (especially air-conditioning) over the past few years has drawn more attention to the residential demand response (RDR) [3]-[4]. In regions with hot summers, high residential electricity consumption has become the main reason for most peak demand and critical peak-valley differences during heat waves [5].

Compared with the commercial demand response (CDR) and industrial demand response (IDR), the main advantages and challenges for RDR are as follows. RDR is less likely to affect the daily production routine and work efficiency, which gains a valuable advantage in DR events. However, the flexibility and uncertainty of the residential electricity consumption pattern complicate the modeling and implementation of RDR. Retailers or aggregators may be required to bring the small-scale flexible RDR to a large-scale market [6].

Given this context, high potential and challenges lie in the electricity market with residential customers as DR providers. Therefore, we study the interactions and corresponding consequences among residential customers, retailers, generation companies (GenCos), and the independent system operator (ISO) in a day-ahead electricity market (DAM) with RDR.

B. Literature Review

RDR has been studied in the existing literature. Muratori et al. [7] evaluate the RDR potential under time-varying pricing by a dynamic energy management framework. Similarly, Morales et al. [8] propose a bilevel model for retailers and residential customers to evaluate the influence of dynamic pricing schemes on the retailers' cost. Nijhuis et al. [9] analyze two different DR programs for residential customers—one from the energy supplier based on the electricity price, and the other from the network operator based on the network load. For those residential customers faced with a flat rate, Zhong et al. [10] develop a coupon incentive-based DR scheme to introduce them to DR events. However, in these studies, the residential customers act on DR issues in a reactive way, where they react to the price or other incentive-based DR signals. Therefore, their movements naturally lag when market situations change.

Several research papers study the proactive RDR and its consequences in the electricity market. Some assess the interactions between retailers and the market operator. Song et al. [11] propose a stochastic programming approach to determine the short-term optimal bidding strategies for retailers in the DAM. Herranz et al. [12] propose a robust methodology for a retailer to determine the optimal bidding strategy. In addition, some focus on the interactions between an aggregator/retailer and its residential customers. Adika and Wang [13] propose a day-ahead demand-side bidding approach for residential customers and the retailer. During the time slot that the demand exceeds the supply, the residential customers bid for the amount they can shed and the compensation they will gain to achieve balance between the power supply and demand. Zhang et al. [14] propose a real-time trading model to optimize the RDR behaviors for a retailer and its consumers. Furthermore, some address the interaction among multiple utility companies/retailers and multiple customers. Kamyab et al. [15] model the DR problem as two noncooperative games. The first game models the supply function bidding mechanism for the utility companies, and the second game models the optimal load shifting strategies for the price anticipating customers. Gkatzikis et al. [16] introduce a hierarchical market model to the aggregators and residential customers, which considered the customers' tradeoff between DR compensation and load shifting discomfort. Moreover, the agent-based modeling and simulation (ABMS) approach may capture the diverse objectives of the multiple entities in the DAM [17]. Bruninx et al. [6] study the strategic interactions between an aggregator, its residential consumers (as DR providers), and the DAM in a bilevel optimization framework. Based on three types of agents, the aggregator-consumer interaction is captured either as a Stackelberg or a Nash Bargaining Game. The results prove that the DR provider-aggregator cooperation may yield critical financial benefits. Similarly, Wang et al. [18] present a multi-agent framework to evaluate optimal RDR implementation via interactions between households and the retailer.

However, the complex interactions among GenCos, retailers, residential customers, and the ISO have not yet been studied. This may lead to underestimation of the RDR potential. Besides, although reinforcement learning (RL) approaches are widely utilized to optimize bidding strategies, the RL training situation that falls in a local optimum solution is rarely discussed in these studies [19]-[21]. In addition, it is significant to find target customers during RDR events. As statistical residential electricity consumption surveys usually address the relationships between consumption patterns and household types, the household types should be taken into consideration in the RDR models [22]-[24].

In general, although RDR has been studied in many papers, the interactions and corresponding consequences between the GenCos, retailers, and residential consumers should be further studied. Moreover, the local optimization possibility of the RL approach should be analyzed and discussed. In addition, the household types should be taken into consideration in the RDR model for heterogeneous residential customers.

C. Innovations and Contributions

Consequently, we study the proactive RDR and interactions among GenCos, retailers, residential customers, and the ISO in a DAM. All entities are assumed to be rational and self-interested. The ISO is responsible for market balance and clearance to maximize social welfare. The GenCos seek optimal bidding strategies to maximize profits and act as utility suppliers. The retailers bridge the gap between the demand side and supply side, and they also seek optimal bidding strategies to maximize profits. The residential customers participate in the DR events by adjusting their consumption patterns according to retail prices.

In general, the main contributions of this paper are as follows:

1) We propose a comprehensive market frame in which residential customers can provide proactive RDR actions. In this novel market frame, we evaluate the corresponding consequences of the interactions among multi-entities via an ABMS approach. These entities are represented by the consumer agents (CAs), retailer agents (RAs), GenCo agents (GAs), and the ISO, respectively.
2) We develop a modified RL procedure for the RAs and GAs to optimize their bidding strategies. This procedure can improve the benefits for the competitive entities in the DAM. Besides, we modify the RL approach to avoid the training result falling into the local optimum solution.
3) We model residential electricity consumption patterns separately according to household types via a bottom-up method. Through this heterogeneous household electricity consumption model, CAs may capture the RDR potential difference among households in advance. Therefore, residential customers may take automatic DR actions via CAs in the DAM.
4) We verified the proposed models and methods based on realistic residential data in China. Moreover, we also compare the proposed market frame and RL procedure with existing approaches via four scenarios. We simulate and analyze the scenarios with/without proactive entities, the modified/conventional RL, and RDR behaviors of different types of households. The simulation results of the interactive market with RDR may provide some information for retailers, regulators, and policymakers.

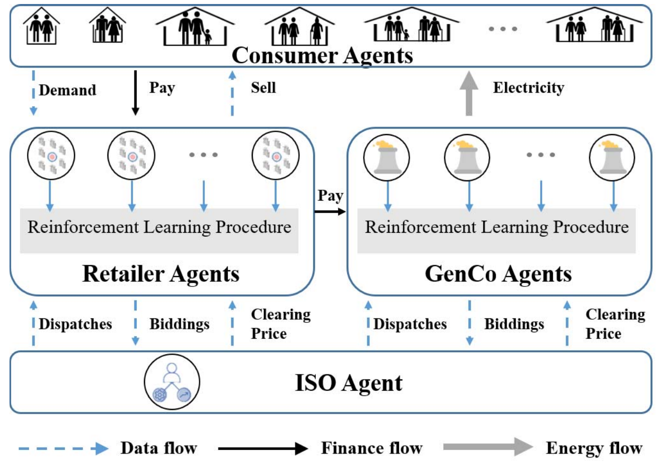

Fig. 1. ABMS framework for the electricity market with RDR.

D. Paper Organization

The remainder of this paper is organized as follows. Section II illustrates the ABMS framework for the electricity market with RDR. Section III constructs the formulation of the problem, including a bottom-up electricity consumption model for residential households of different household structure types, and a multi-agent model for the electricity market with RDR. Section IV presents the modified RL procedure. Section V describes numerical examples based on a pilot project carried out in China. Section VI discusses the conclusions and future work.

II. Framework for Electricity Market with Residential Demand Response

The framework for the electricity market with RDR is shown in Fig. 1, which includes four parts, the ISO, GAs, RAs, and CAs. They represent the market operator, suppliers, retailers, and consumers, respectively. There are three flows in this framework. The data flow allows different agents to communicate with each other using IoT technologies. It carries many types of information, including the hourly electricity demand profiles from the CAs, the bidding information from the RAs and GAs, along with the hourly dispatches and market clearing prices from the ISO. The finance flow is among the customers, retailers, and GenCos. The energy flow is between the GenCos and customers.

To capture the influence of RDR on the electricity market,

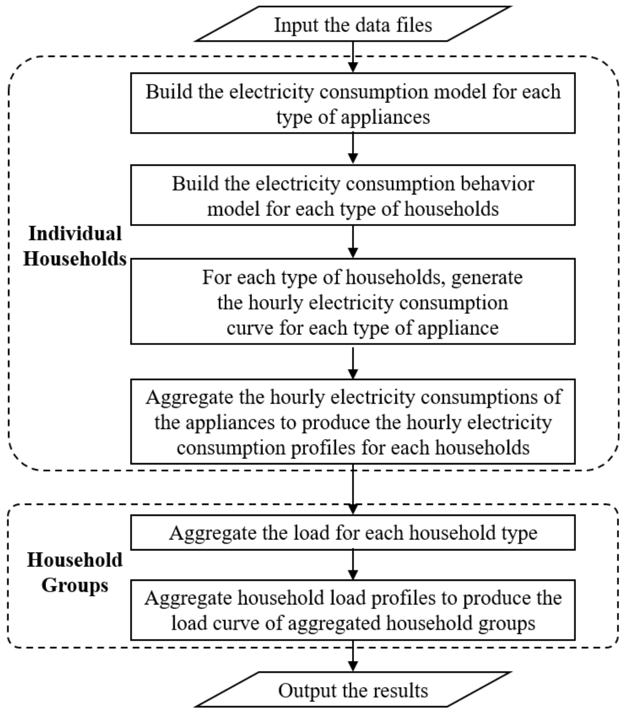

Fig. 2. Bottom-up modeling procedure.

the first step is to model the residential electricity consumption patterns. The bottom-up approach can describe the hourly consumption detail of each appliance. Moreover, to take RDR separately by different types of households and appliances. The second step is to model all agents. This framework is designed to simulate the interactions of different entities and analyze the consequences of these interactions. The CAs send the predicted next-day demand to the RAs. The RAs play an active role by bidding with GAs. The ISO clears the market to maximize social welfare. The CAs reschedule the electricity consumption and resubmit the new demand to the system. The interactive calculations end when the difference between the last two market clearing prices is smaller than ε .

III. Problem Formulation

A. Residential Electricity Consumption Model

Led by the CA, the appliances of each residential customer are monitored and measured by the HEMS. The consumption behavior differs between households, which leads to the RDR potential difference. So the residential electricity consumption model specifically considers the household factor to capture the consumption difference among different types of households. In this paper, the bottom-up approach is used to model the residential electricity consumption behavior, as it can analyze each individual appliance's effect on the total load, which is necessary in the RDR simulation process.

As shown in Fig. 2, the first step of the bottom-up model is to model the appliances individually. Then, the consumption model of each household is established. Lastly, the consumption model of the household groups is established [25]-[26].

1) Electricity Consumption Model for the Individual Household

Common end-use appliances, such as air conditioners, electric water heaters, TV sets, and washing machines are modeled in this step. The air conditioner load is modeled separately from the other appliances because of its large share of the residential electricity load and its thermally controlled features [27]. One single day is separated into 24 time intervals, for example, 1:00-2:00 is the first time interval, $t=1$, 2:00-3:00 is the second, $t=2$, and so on. Then the residential consumption is modeled for each time interval. In each interval, it should be noted that the load is modeled by the time resolution of one minute.

Consumption model for each appliance: For household h of household type n , the consumption model of appliance k and air conditioner v are modeled as follows, respectively [28]-[29]:

$$\begin{aligned}
&L_{n,h,k}(t)=P_{n,h,k}\cdot\Gamma_{n,h,k}(t)\\
&L_{n,h,v}(t)=\gamma_n\left[T_{out}(t)-T_{n,h,set}(t)\right]\cdot P_{n,h,v}\cdot O_{n,h,v}(t)\\
&0<\gamma_{n,h,v}<1,\qquad n=1,2,\ldots,N\\
&h=1,2,\ldots,H_n,\quad k=1,2,\ldots,K_n\\
&v=1,2,\ldots,V_n,\quad t=1,2,\ldots,24
\end{aligned}\tag{1}$$

where $L_{n,h,k}(t)$ is the energy consumption of appliance k , household h from household type n . There are H_n individual households for household type n . There are N types of household in this paper. Except for air conditioners, there are K_n appliances for each household of household type n . $P_{n,h,k}$ is the fixed power of appliance k . $\Gamma_{n,h,k}(t)$ is the total operating time of appliance k , where the time resolution is one minute. $\Gamma_{n,h,k}(t)=0$ if appliance k does not work during time interval t . $L_{n,h,v}(t)$ is the energy consumption of air conditioner v . There are V_n air conditioners in total per household. $\gamma_{n,h,v}$ is the air conditioner refrigeration efficiency parameter. $T_{out}(t)$ is the outside temperature. $T_{n,h,set}(t)$ is the cooling set temperature. $P_{n,h,v}$ is the fixed power of the air conditioner. $O_{n,h,v}(t)$ is the ON/OFF variable of air conditioner v .

Consumption model for each household: Therefore, for household h of household type n , the electricity consumption is:

$$L_{n,h}(t)=\sum_{k=1}^{K_n}L_{n,h,k}(t)+\sum_{v=1}^{V_n}L_{n,h,v}(t)\tag{2}$$

where $L_{n,h}(t)$ is the total electricity consumption for household h of type n .

2) Electricity Consumption Model for Household Groups

The last step is to model the electricity consumption of household groups, as is shown in Fig. 2:

$$L(t)=\sum_{n=1}^{N}\sum_{h=1}^{H_n}L_{n,h}(t)\tag{3}$$

where $L(t)$ is the total electricity consumption for a residential group with N types of households.

To simplify the expression, the rest part only describes the formulation during one-time interval. These approaches are also suitable for other time intervals.

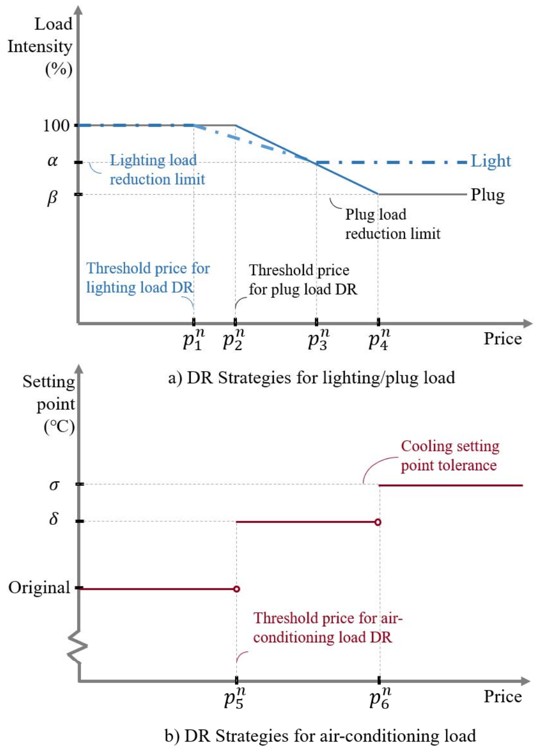

Fig. 3. DR strategies for CAs.

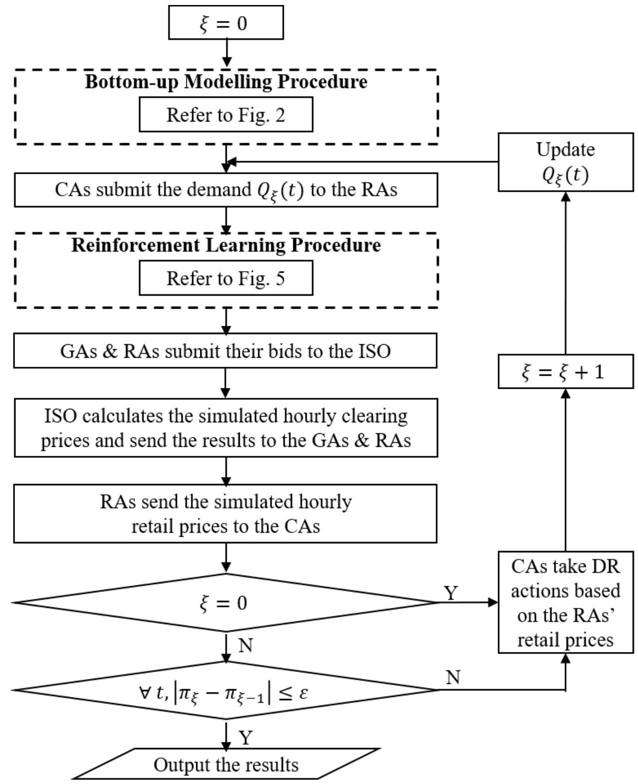

Fig. 4. General process for the agents.

B. Agent-based Model for the Electricity Market with RDR

1) RDR Model for CAs:

As mentioned above, CAs have two functions: Forecast the hourly electricity demand of the residential consumers using the bottom-up model mentioned above and take DR actions according to the retail prices of the retailers. They also model and simulate the RDR strategies according to retail prices.

If the CAs operate randomly, shutting down or rescheduling one or two appliances will be disastrous for the residents. To achieve financial profits from DR actions and avoid disturbing the original residential living habits, the household appliances are separated into different groups by CAs. Group 1 consists of the lighting appliances, which are necessary during the nighttime. Group 2 consists of the non-emergency appliances, referred to as the "plug load" in this paper. They are the washing machines, laptops, electric water heaters, electric cookers (rice cookers, electric kettles, and microwave ovens), and refrigerators. Group 3 consists of the air conditioners.

Considering the end users' comfort factor, some constraints are made for the CAs during DR events. Since customers need lighting loads during the nighttime, the Group 1 appliances are only allowed to take DR actions during the daytime. Additionally, the adjusting temperature for Group 3 is limited to be within two degrees Celsius. The DR strategies for the plug load and the lighting load are shown in Fig. 3 a). The DR strategies for air conditioners are shown in Fig. 3 b).

Assuming the CAs will be triggered to take actions when the electricity price reaches a certain threshold, p_1^n, p_2^n, and p_5^n are the threshold prices for lighting, plug, and air-conditioning loads of household structure type n, respectively. α, β, and σ are the load reduction limits for lighting, plug, and air-conditioning loads, respectively. The limits will occur when the price exceeds p_3^n, p_4^n, and p_6^n, respectively.

2) Bidding Model for RAs:

In this DAM simulation model, RAs actively submit their bid utility functions to the ISO. Compared with the GAs who intend to sell electricity at a higher price to earn more profit, RAs seek ways to purchase more utilities at a lower price to gain more profit. In this paper, the strategic bidding behavior for RAs is described by the markdown rates. Therefore, the bid utility function of RAs can be described as:

$$\begin{gathered}\tilde{S}_r\left(w_r\right)=\varphi_r \mu_r w_r^2+\phi_r \mu_r w_r+\omega_r, \\ r \in\{1,2, \ldots, R\},\ w_r \in\left[w_r^{\min }, w_r^{\max }\right],\ \mu_r \in(0,1]\end{gathered} \tag{4}$$

where r represents the RA, and there are R RAs in total. $\tilde{S}_r\left(w_r\right)$ is the bid utility function for RA r under the load demand w_r. φ_r, ϕ_r, ω_r are three parameters. $w_r^{\min}$ is the lower limit and $w_r^{\max}$ is the upper limit, respectively. The markdown rate μ_r is greater than 0 and equal or less than 1. For instance, if a certain RA bids 70% of its actual retail price, then the μ_r is 0.7. When the markdown rate μ_r is 1, that means the RA bids at the real retail price. This real retail price is expressed as $S_r\left(w_r\right)$.

3) Bidding Model for GAs:

Each GA is represented by only one generator. The GA can submit its marginal cost function to the ISO as a bid, which can be the real marginal cost function or an adjusted cost function. Since all the bids are considered to be the "real marginal" cost functions by the ISO, the GAs try to obtain more profits through strategic bidding [19]. The cost function can be defined as:

$$\tilde{C}_g\left(q_g\right) = a_g m_g q_g{}^2 + b_g m_g q_g + c_g, \quad g \in \{1,2,...,G\}, q_g \in \left[q_g^{\min}, q_g^{\max}\right], m_g \geq 1 \tag{5}$$

where g represents the GA, and there are G GAs in total. $\tilde{C}_g\left(q_g\right)$ is the strategic bidding cost function for GA g under electricity production q_g . a_j , b_j , c_j are three parameters. $q_g^{\min}$ is the lower limit and $q_g^{\max}$ is the upper limit, respectively. The markup rate m_g is equal to or greater than 1.0. For instance, if a GA bids at 150% of its real cost, then the markup rate is 1.5. When the markup rate is 1, that means the GA bids at the real cost. The real cost is expressed as $C_g\left(q_g\right)$.

4) Optimization Model for ISO:

The goal of ISO is to regulate the market by maximizing social welfare. Assuming the electricity demand is L , which is sent by CAs. The objective function is:

$$\begin{aligned} &\operatorname{Max}\left[\sum_1^R \tilde{S}_r\left(w_r\right) - \sum_1^G \tilde{C}_g\left(q_g\right)\right] \\ &q_g^{\min} \leq q_g \leq q_g^{\max} \qquad g = 1,2,....,G \\ &w_r^{\min} \leq w_r \leq w_r^{\max} \qquad r = 1,2,...,R \\ &\sum_1^R w_r = \sum_1^G q_g = L \end{aligned} \tag{6}$$

Since the problem is a convex optimization problem, it can be transformed into an equivalent dual problem constrained to the Karush-Kuhn-Tucker (KKT) conditions. So, the dual problem is defined as the Lagrangian [30]:

$$\begin{aligned} \mathscr{L}\left(q,w,\lambda,\beta\right) = &\sum_1^G \tilde{C}_g\left(q_g\right) - \sum_1^R \tilde{S}_r\left(w_r\right) + \\ &\lambda_1\left(L - \sum_1^G q_g\right) + \lambda_2\left(L - \sum_1^R w_r\right) + \\ &\sum_1^G \beta_{1,g}\left(q_g - q_g^{\max}\right) + \sum_1^G \beta_{2,g}\left(q_g^{\min} - q_g\right) + \\ &\sum_1^R \beta_{3,r}\left(w_r - w_r^{\max}\right) + \sum_1^R \beta_{4,r}\left(w_r^{\min} - w_r\right) \end{aligned} \tag{7}$$

where $\lambda_1, \lambda_2, \beta_{1,g}, \beta_{2,g}, \beta_{3,r}, \beta_{4,r}$ are the KKT multipliers. The KKT conditions are:

$$\begin{aligned} &\frac{\partial \mathscr{L}}{\partial q_g} = 0 \qquad \frac{\partial \mathscr{L}}{\partial w_r} = 0 \\ &\lambda_1 \neq 0 \qquad \lambda_2 \neq 0 \\ &\beta_{1,g} \geq 0 \qquad \beta_{2,g} \geq 0 \\ &\beta_{3,r} \geq 0 \qquad \beta_{4,r} \geq 0 \\ &\beta_{1,g}\left(q_g - q_g^{\max}\right) = 0 \qquad \beta_{2,g}\left(q_g^{\min} - q_g\right) = 0 \\ &\beta_{3,r}\left(w_r - w_r^{\max}\right) = 0 \qquad \beta_{4,r}\left(w_r^{\min} - w_r\right) = 0 \\ &L - \sum_1^G q_g = 0 \qquad L - \sum_1^R w_r = 0 \\ &q_g^{\min} \leq q_g \leq q_g^{\max} \qquad w_r^{\min} \leq w_r \leq w_r^{\max} \\ &g = 1,2,\ldots,G \qquad r = 1,2,\ldots,R \end{aligned} \tag{8}$$

According to equl. (7) and (8), we have:

$$\begin{aligned} &\frac{\partial \mathscr{L}}{\partial q_g} = \frac{d\tilde{C}_g\left(q_g\right)}{dq_g} - \lambda_1 + \beta_{1,g} - \beta_{2,g} = 0 \\ &\frac{\partial \mathscr{L}}{\partial w_r} = \frac{d\tilde{S}_r\left(w_r\right)}{dw_r} - \lambda_2 + \beta_{3,r} - \beta_{4,r} = 0 \\ &\forall g = 1,2,\ldots,G \qquad \forall r = 1,2,\ldots,R \end{aligned} \tag{9}$$

The market clearing price π should be equal to or higher than the shadow price. If not, the GAs should purchase energy on the market. So the market clearing price π is:

$$\pi = \max\left[\left(\lambda_1 - \beta_{1,g} + \beta_{2,g}\right), \left(\lambda_2 - \beta_{3,r} + \beta_{4,r}\right)\right] \tag{10}$$

In this dual formulation, the RAs and the GAs must decide their optimal markdown rate μ_r or markup rate m_g , which is solved via the next RL procedure. The RL procedure helps GAs or RAs choose the optimal markup rate m_g or markdown rate μ_r during the bidding iteration.

5) General Process for the Multi-agent System:

The general simulation process is shown in Fig. 4. In the DAM, first, the CAs forecast the second-day residential hourly electricity demand using the bottom-up modeling procedure. Then, the CAs submit the original demand to the RAs. Before the deadline, the RAs and GAs submit their bidding for the next day to the ISO via the RL procedure which is described in section IV. After that, the ISO calculates the simulated hourly clearing prices. Then, the RAs send the simulated hourly retail prices to the CAs. Based on the retail prices, the CAs update the residential electricity demand. All agents repeat the steps iteratively until the difference of the last two clearing prices is smaller than ε . After that, the real market clearing results for the next day are announced by the ISO. Finally, all the agents take the results and adjust their next-day supplies or demand schedule. It should be noted that only the final market clearing results are officially announced. The others are just the intermediate calculation process.

IV. REINFORCEMENT LEARNING PROCEDURE

RL plays a critical role in the machine learning theory. The other two important categories of machine learning are supervised learning and unsupervised learning. Compared with other machine learning methods, the learning model of RL is very similar to the learning process of human beings [31]. The agents involved in the RL procedure interactively learn from the environment according to the corresponding reward of each action and output the optimal decision at the end of the training procedure. Supervised and unsupervised learning are more suitable for problems like cluster analysis [32]. So, in this paper, we use RL to solve the problem. During the training process, agents renew their action-policy according to the reward they gain in the previous round. To simplify the expression, in this RL model, both GAs and RAs are marked as agent j , $j \in \left[1, G+R\right]$. Suppose there are D training iterations in the RL procedure. Before the iterations starts, for agent j, there are θ_j different markup or markdown rates for it to choose from, and all the choices form the action space A_j . In iteration d, agent j takes action i to win the reward by the simulation system. The

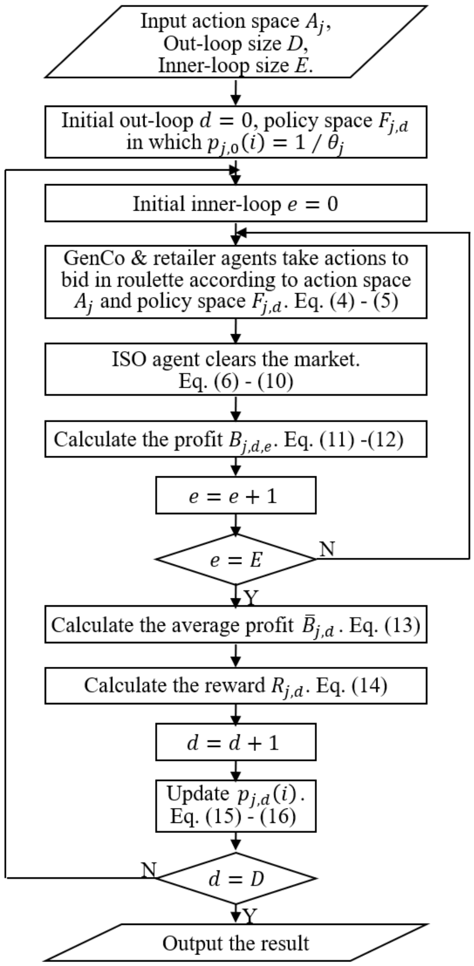

Fig. 5. Reinforcement learning procedure.

corresponding selection probability of action i for agent j is $p_{j,d}(i)$, and they form the policy space $F_{j,d}$. The policy space is renewed every training by the reward of the action. The agent takes action according to the new policy at the next training step $d+1$. The iteration keeps going and the policy space gradually converges. Eventually, the RL procedure will end and produce the final optimal policy space.

However, in this formulation, the policy may be influenced by one reward and fall into a local optimum solution. To prevent this, we change the procedure into a nested loop, as shown in Fig.5. The outer loop is still counted by d, and the inner loop is counted by e. The total iteration numbers for the outer loop and inner loop are D and E, respectively. In the inner loop e, all the agents take actions to bid in roulette according to the action space A_j and policy space. $F_{j,d}$. The reward for action i in the inner loop d,e is the difference between the profit from this action and the average of the whole inner loop. In this way, the reward is not only decided by a single action but is also related to all the other actions in the entire inner loop. This will reduce the impact of a single exploration on the entire training process.

The profits for each agent can be expressed by equations (11) and (12). Equation (11) illustrates the profits for GAs, and equation (12) illustrates the profits for RAs.

$$B_{j,d,e} = q^*_{j,d,e}\left[\pi^*_{d,e} - C_j(q^*_{j,d,e})\right],\ j \in \{1,2,...,G\} \tag{11}$$

where $B_{j,d,e}$ is the profit for GA j, when $j \in \{1,2,...,G\}$; $q^*_{j,d,e}$ is the assigned dispatch to GA j, and $C_j(q^*_{j,d,e})$ is the real cost for GA j; $\pi^*_{d,e}$ is the market clearing price.

$$B_{j,d,e} = w^*_{j,d,e}\left[S_j(w^*_{j,d,e}) - \pi^*_{d,e}\right], j \in \{G+1,...,G+R\} \tag{12}$$

where $B_{j,d,e}$ is the profit for RA j, when $j \in \{G+1,...,G+R\}$; $w^*_{j,d,e}$ is the assigned dispatch to RA j, and $S_j(w^*_{j,d,e})$ is the real retail price for RA j; $\pi^*_{d,e}$ is the market clearing price.

The average profit for agent j in this inner loop is:

$$\bar{B}_{j,d} = \frac{1}{E}\sum_{e=1}^{E} B_{j,d,e} \tag{13}$$

Then the corresponding reward for agent j is:

$$R_{j,d} = B_{j,d,E} - \bar{B}_{j,d} \tag{14}$$

As mentioned above, instead of using the profits as the reward function, this method can describe the progress better and ensure learning is headed in the right direction. When the $R_{j,d}$ is negative, the profit of this action is below the average and not worth trying again.

The outer loop uses the reward to renew the $p_{j,d+1}(i)$. The iterative process gradually converges to the optimal solution, and the corresponding probability of this solution is close to 1. Suppose agent j takes action s in outer loop d. The renewed selection probabilities are:

$$\tilde{p}_{j,d+1}(i) = \begin{cases} \max\left[\left(\rho\cdot\left|R_{j,d}\right|\cdot p_{j,d}(i) + R_{j,d}\right), 0\right], & i = s \\ \max\left[\rho\cdot\left|R_{j,d}\right|\cdot p_{j,d}(i), 0\right], & i \neq s \end{cases} \tag{15}$$

where $\tilde{p}_{j,d+1}(i)$ is the intermediate variable and ρ is the learning intensity parameter.

The sum of all probabilities should be 1. Therefore, the calculation results should be standardized:

$$p_{j,d+1}(i) = \begin{cases} \dfrac{\tilde{p}_{j,d+1}(i)}{\sum \tilde{p}_{j,d+1}(i)} & \text{if } \sum \tilde{p}_{j,d+1}(i) \neq 0 \\ p_{j,d}(i) & \text{if } \sum \tilde{p}_{j,d+1}(i) = 0 \end{cases} \tag{16}$$

It should be noted that when the market environment is rather stable, the rules and agents stay the same, then it is unnecessary to re-run the training procedure every single day. This RL training is well adapted to the market environment, and it can be renewed intermittently. The strategy can be revised and fine-tuned as subtle changes occur in the marketplace.

V. Numerical Analysis

We simulate and analyze the interactions among multi-entities and their corresponding consequences in the proposed DAM via a comprehensive large case, in which 30,000 households are assumed to provide proactive RDR. According to the previous survey, these households can be divided into three types—Nuclear Family as Type A, Retired Elderly Family as Type B, and Extended Family as Type C. Suppose two RAs aggregate residential loads and bid together with five GAs in the DAM. The ISO takes the bidding information from these RAs and GAs and clears the market hourly. In detail, common household appliances' information is shown in Table I, where the average possession rate means the probability of holding

TABLE I
COMMON HOUSEHOLD APPLIANCES

Type	Average possession rate	Fixed power (W)	Standby Power (W)	Average working frequency (per day)	Average working cycle (min)
Microwave Oven	0.62	1500	NA	5	5
Refrigerator	0.89	110	8.1	40.5	12
Electric Kettle	0.97	1500	NA	4	12
Rice Cooker	1	2000	NA	2	35
Wash Machine	0.70	1200	NA	1	25
TV	0.78	200	4	2.62	40
Computer	0.59	110	2.5	2.5	60
Air Conditioner	0.95	1300	NA	3.36	150
Water Heater	0.79	1600	NA	3	20
Light	1	120	NA	5	90

such an appliance. Table II indicates the proportion of different types of households. The parameters for GAs and RAs are shown in Tables III and IV, respectively. The numerical case is carried out by MATLAB.

A. Electricity Consumption for Residential Households

These simulations are based on realistic residential data from a city in the eastern coastal region of China [26]. One summer working day is chosen as residential peak load usually comes with heat waves [33]. The outside temperature is shown in Fig. 6.

The hourly electricity consumption results of different types of households according to the bottom-up model are shown in Figs. 7-10. The cookers in Figs. 7-9 include the rice cooker, microwave oven, and electric kettle. As shown in Fig. 7, the load for Type A households remains low during 9:00-16:00 as most of the occupants of household Type A are typically at work or school. By contrast, in Figs. 8 and 9, the loads of households which contain elderly citizens (Types B & C) show no obvious valleys as the retired elderly typically stay at home during that period. Households of Type B show a small consumption valley during 18:00-20:00, as the elderly in China have the habit of having a long walk after their dinner which they believe will benefit their health. Generally, the cookers, water heaters, and air conditioners consume most of the electricity. According to Fig. 10, because the population of Type C is larger compared to Types A & B, they consume more electricity, especially during peak hours.

B. Reinforcement Learning Procedure for GAs & RAs

After the bottom-up procedure, the GAs and RAs are involved in the RL procedure. The CAs will not participate in this step, which means RDR actions are temporarily ignored in this part. To compare the proposed methods with the existing techniques, three different scenarios are analyzed, and the results are shown in Fig. 11.

Scenario 1: Without RAs, GAs, and CAs

This scenario simulates the market without interactive GAs and RAs. Suppose each GenCo or retailer submits the same bidding function for 24 intervals. In Fig. 11, the solid black line with squares indicates the market clearing price for this scenario.

Scenario 2: GAs & RAs with a Conventional RL Procedure, without CAs

In this scenario, the interactive GAs and RAs use conventional RL to determine their bidding strategies. The conventional RL does not consider the possibility of falling into local optimum. Compared with the proposed modified RL procedure, the conventional RL procedure lack of an inner loop to moderate the influence of one single training. In Fig. 11, the

TABLE II
HOUSEHOLD TYPES AND THEIR RATES

Type	Number of Residents	Characteristics	Proportion (%)
A	3	Nuclear family	58
B	2	Retired elderly couples	19
C	4~5	Extended family	23
Total	---	---	100

TABLE III
PARAMETERS FOR GAS

GAs	a_g	b_g	$P_{\min}$ (MW)	$P_{\max}$ (MW)
G_1	0.04	2.00	0	25
G_2	0.35	2.00	0	25
G_3	0.04	1.80	0	25
G_4	0.02	1.50	0	40
G_5	0.04	2.00	0	30

TABLE IV
PARAMETERS FOR RAS

RAs	φ_r	ϕ_r	$W_{\min}$ (MW)	$W_{\max}$ (MW)
R_1	0.10	30.00	0	50
R_2	0.08	25.00	0	60

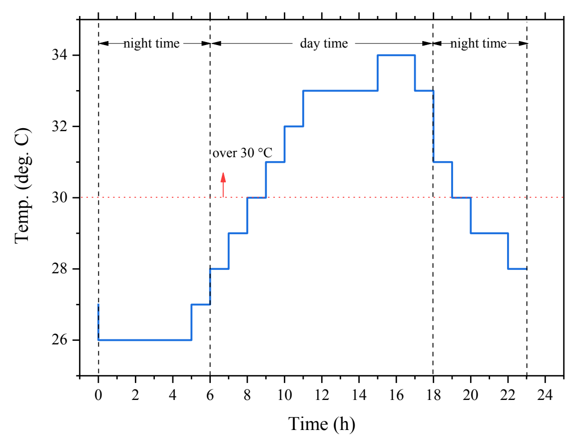

Fig. 6. Outdoor temperature in a certain summer day.

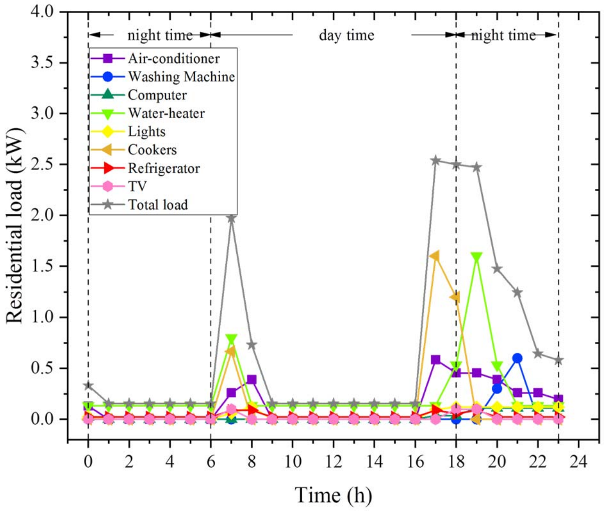

Fig. 7. Hourly load for the household Type A.

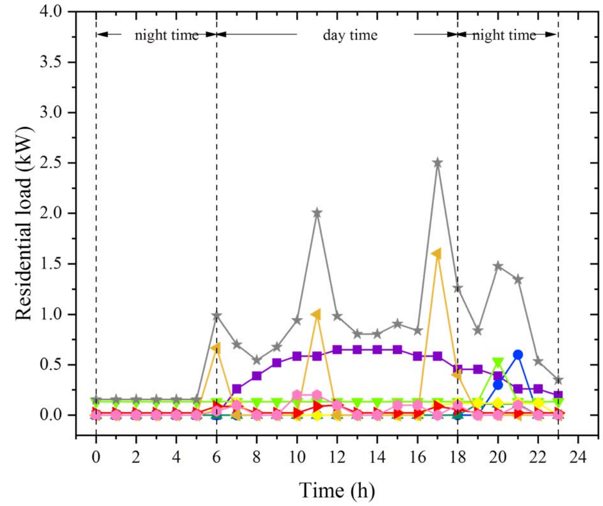

Fig. 8. Hourly load for the household Type B.

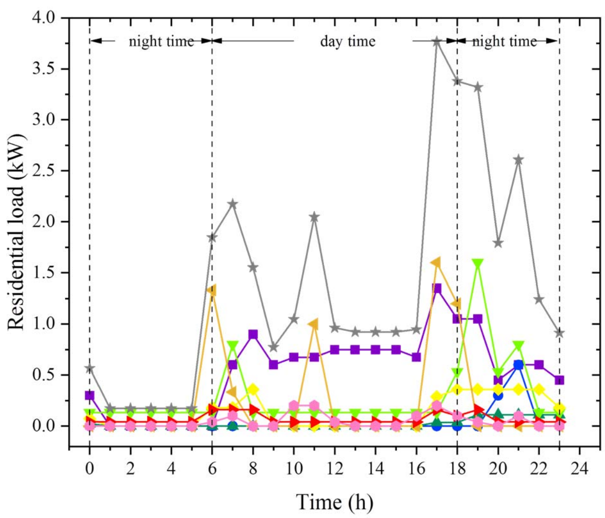

Fig. 9. Hourly load for the household Type C.

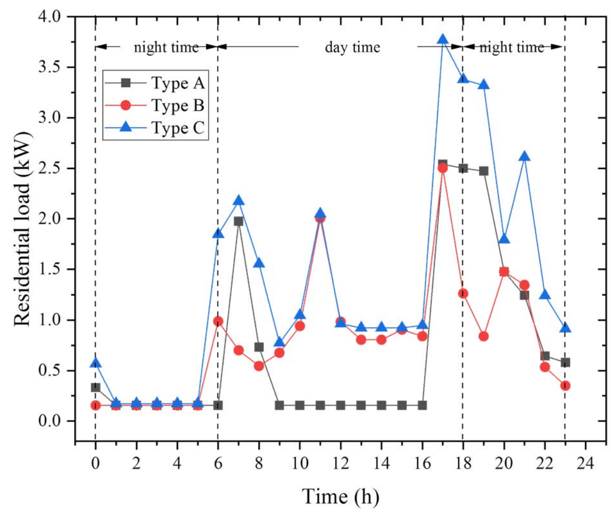

Fig. 10. The comparison of hourly load.

blue dot-dash line with triangles indicates the market clearing price under this scenario.

Scenario 3: GAs & RAs with Modified RL Procedure, without CAs

The interactive GAs and RAs use modified RL to determine their bidding strategies in this scenario. The red dash line with circles in Fig. 11 indicates the market clearing price under this scenario.

As shown in Fig. 11, the market clearing prices of different scenarios are obviously different from each other. So, it can be inferred that the iterations based on RL affect the bidding decisions of the GAs and RAs.

In scenario 1, neither the GAs nor RAs are trained. It outputs the highest curve of market clearing prices among the three scenarios. Scenarios 2 & 3 with interactive RAs and GAs output lower market clearing prices than scenario 1. Therefore, the interactions between RAs, GAs and the ISO may improve social welfare. The modified RL procedure in scenario 3 reduces the market clearing prices especially during peak hours.

Compared with scenario 3, the conventional RL procedure in scenario 2 is not sensitive enough to the changes in electricity demand. As highlighted by a grey circle in Fig. 11, although the load is lower at the 18:00 interval compared with the 17:00 interval, the agents in scenario 2 keep bidding higher offers at the 18:00 interval, which leads to the higher market clearing price. This unsuccessful bidding behavior stop after the 19:00 interval, which is one step behind the load change. At the 17:00 interval, the price gap between scenario 1 and scenario 3 reaches 0.74 CNY/kWh.

The market behaviors of GAs differ from the time intervals and gradually change by the number of iterations. In using G_4 as an example, in time interval 17:00, the convergence of the conventional RL and the modified RL are as shown in Fig. 13. For the conventional RL, the reward curve converges when the number of training iterations reached 3500 (Fig. 12). For the modified RL, the reward curve converges when the number of training iterations reached 2000 (Fig. 13). The modified RL takes less training iterations to converge.

In the modified RL procedure scenario, the markup rate of G_4 becomes 1.5 during time intervals 7:00, 17:00, 18:00, and 19:00, and remains at 1.2 during the other time intervals. It can be inferred that G_4 intends to gain more profit by bidding for higher price during the peak load time intervals.

The proposed modified RL procedure in this paper shows advantages in fast convergence and robust adaptation to

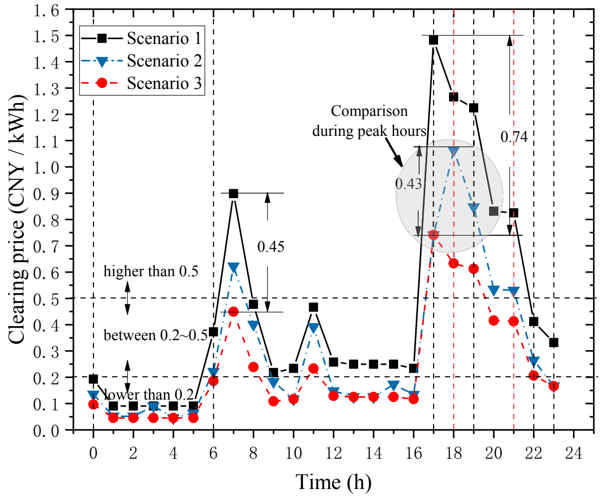

Fig. 11. Clearing price comparison of three scenarios.

environmental changes.

C. Demand Response Behavior of CAs

1) Analysis for the ABMS Results:

In this part, we compare the ABMS results between scenario 3 and scenario 4. Scenario 4 adds the CAs with RDR on the basis of scenario 3.

Scenario 4: GAs & RAs with Modified RL Procedure and CAs with RDR

In Fig. 14, the solid black line with squares shows the hourly clearing prices of scenario 3, and the solid red line with circles shows the hourly clearing prices of scenario 4. Compared to the price without RDR (scenario 3), the price with RDR (scenario 4) drops by 0.12 Chinese Yuan (CNY) at the 7:00 interval and 0.16 CNY at the 17:00 interval. In scenario 4, the intervals with price lower than 0.2 CNY stay nearly the same as scenario 3,

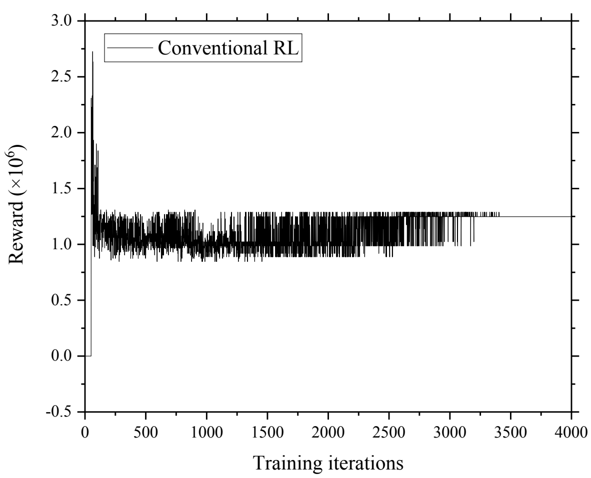

Fig. 12. Conventional RL Training Iterations.

but the intervals where price is higher than 0.5 CNY are reduced by two hours. Above all, the price curve becomes smoother after the residential consumers take proactive DR actions via CAs.

Table V shows the detailed comparison of the simulation results of four different scenarios. Since the RDR behavior is only studied in scenario 4, the residential load remains the same

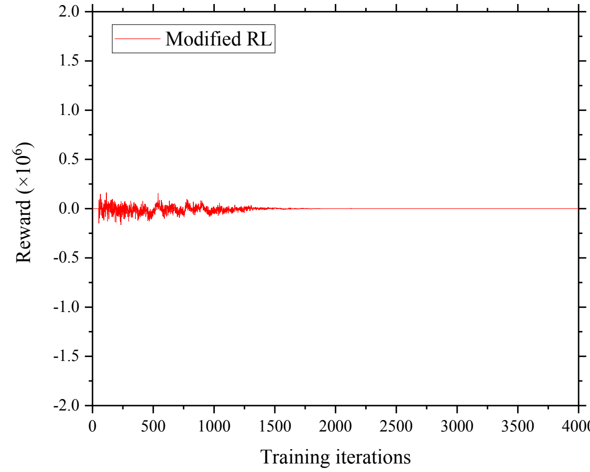

Fig. 13. Modified RL Training Iterations.

for the first three scenarios. The lowest clearing price stays the same for scenarios 2, 3, and 4, while the highest price and the price peak-valley difference differ from each other.

The comparison between scenario 1 and the other ABMS market simulation results indicates that the market without interactive entities provides the worst results, in which the clearing price and the price peak-valley difference is the highest among all four scenarios. The interactions between GAs, RAs and the ISO optimize the electricity market clearing prices in scenarios 2, 3, and 4. In scenario 3, the agents using the conventional RL procedure reduce the peak-valley price difference from 1.392 to 1.018 (CNY / kWh), which is by around 27%. In scenario 4, the agents using the modified RL procedure output better results. They reduce the peak-valley price difference from 1.392 to 0.696 (CNY / kWh), which is by around 50%. Compared with the other scenarios, scenario 4 provides the best results. The peak-valley price difference is

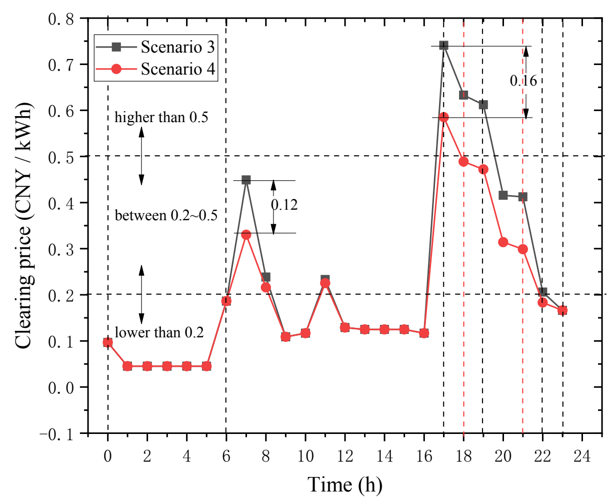

Fig. 14. Clearing price with/without RDR.

0.54 (CNY / kWh), which is around 61% less than the scenario 1.

2) Analysis for the RDR Results

As shown in Fig. 15, the solid black line with squares is the total residential load demand without RDR, while the green dot-

TABLE V
COMPARISON OF FOUR DIFFERENT SCENARIOS

Comparison		Scenario 1	Scenario 2	Scenario 3	Scenario 4
Clearing Price (CNY / kWh)	Highest	1.482	1.063	0.741	0.585
	Lowest	0.090	0.045	0.045	0.045
	Peak-Valley Difference	1.392	1.018	0.696	0.540
Residential Load (MW)	Highest	83.057	83.057	83.057	65.518
	Lowest	4.784	4.784	4.784	4.784
	Peak-Valley Difference	78.273	78.273	78.273	60.734

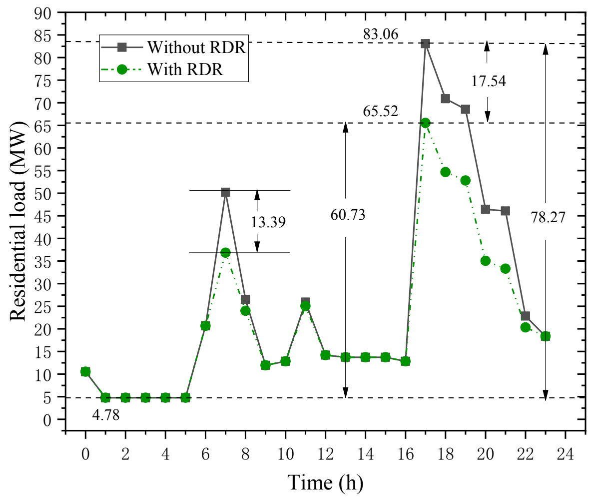

Fig. 15. Total residential load demand with/without RDR.

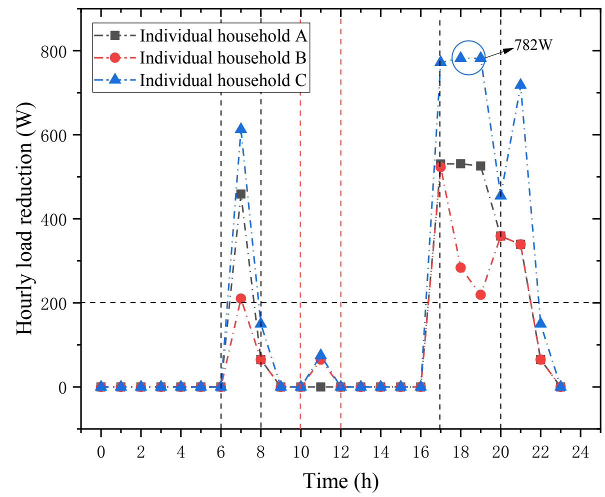

Fig. 16. Hourly load reduction of individual household.

dash line with circles is the total residential load demand with RDR. Compared with the residential load without RDR, the peak load with RDR drops from 83.06 MW to 65.52 MW, while the valley stays the same at 4.78 MW. Meanwhile, the peak-valley difference transforms from 78.27 MW to 60.73 MW. Therefore, the peak-valley difference is reduced by 17.54 MW, which is around 22% lower than before. In addition, the peak load during the daytime is reduced by 13.39 MW, as well. The households of Type C save the most of money as they reduce most of the loads.

Fig. 16 shows the hourly load reduction of the individual household. The black dot-dash line with squares is the hourly load reduction of the individual Type A household, while the red dot-dash line with circles is the hourly load reduction of the individual Type B household, and the blue dot-dash line with triangles is the hourly load reduction of the individual Type C household. The load reduction of each type of household over 200 W happens at 7:00 and during 17:00-21:00. The household of Type C reduces the most load during the entire day, and the highest load reduction happens during the 18:00 and 19:00 intervals, which is 782 W. During 7:00 and 17:00-19:00 intervals, the Type A household reduces more load than the Type B household. However, during the 11:00 interval, the Type B household reduces more load than the Type A household. It indicates that the reduction is related to the number of occupants and their consumption habits. The Type C households have more occupants than Types A or B, which leads to higher energy consumption and higher potential in RDR. Although the Type B households have fewer occupants–retired elderly citizens, it reduces more load than the household of Type A during the 11:00 interval, because the occupants of

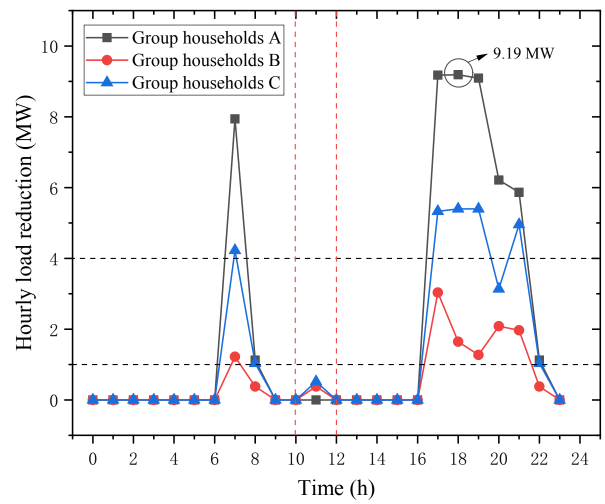

Fig. 17. Hourly load reduction of household groups.

Type A are not at home while the elderly of Type B cook lunch for themselves during that time. Moreover, the load recovery may cause a "re-peak" after the peak load periods. As household Type C reduces the most load during the peak hours, it is highly possible for these households to rebound after the peak load periods. Fig. 17 shows the hourly load reduction of household groups. Dramatically, the group that reduces the most load for almost the entire day is group A rather than group C. This is because Type A households take the majority (over 50%) of all households.

VI. CONCLUSIONS AND FUTURE WORK

This paper models and simulates the proactive RDR electricity market consisting of multiple self-interested and non-cooperative entities using the ABMS method.

The results show that proactive RDR programs may stimulate underlying RDR potentials, and eventually bring more benefits to the customers and improve power system efficiency and social welfare. Moreover, the interactions among proactive market entities (GenCos, retailers, and residential customers) can reduce the clearing prices, especially during peak load hours, which may yield significant financial benefits for both the supply and demand sides. Furthermore, the RL process improves the bidding decisions for the retailers and GenCos. The proposed modified RL is more robust and takes less training steps to execute and converge. In addition, the RDR via CAs narrows the peak-valley difference by 22%. Besides, the RDR potential differs with different types of households, which is related to the consumption pattern. Households of the extended family type possess the highest RDR potential amongst all three household types in this paper.

GenCos, retailers, market operators, and policymakers can use the proposed methods and models to evaluate the interactions among multiple market entities. They can also use the proposed models to assess RDR potentials of different households so as to locate target customers.

In future work, we will analyze the impact of network constraints on the interactive market frame and study the detailed model for load recovery after peak hours.

REFERENCES

[1] P. Siano, "Demand response and smart grids — A survey", *Renewable and Sustainable Energy Reviews,* vol. 30, pp. 461-78, 2014.

[2] J. Zhang, P. Zhang, H. Wu, et al., "Two-stage load-scheduling model for the incentive-based demand response of industrial users considering load aggregators", *IET Generation, Transmission & Distribution*, vol. 12, pp. 3518-26, 2018.

[3] National Energy Administration of China, "A 8.5% growth in electricity consumption of the whole society in 2018", http://www.nea.gov.cn/2019-01/18/c_137754978.htm

[4] D. Xie, H. H. Hui, Y. Ding, et al., "Operating reserve capacity evaluation of aggregated heterogeneous TCLs with price signals", *Applied Energy*, vol. 216, pp. 38-47, 2018.

[5] S. Mhanna, A. C. Chapman and G. Verbič, "A Fast Distributed Algorithm for Large-Scale Demand Response Aggregation," in IEEE Transactions on Smart Grid, vol. 7, no. 4, pp. 2094-2107, July 2016.

[6] K. Bruninx, H. Pandžić, H. Le Cadre and E. Delarue, "On the Interaction Between Aggregators, Electricity Markets and Residential Demand Response Providers," in IEEE Transactions on Power Systems, vol. 35, no. 2, pp. 840-853, March 2020.

[7] M. Muratori, G. Rizzoni, "Residential demand response: Dynamic energy management and time-varying electricity pricing", *IEEE Transactions on Power Systems*, vol. 31, pp. 1108-17, 2016.

[8] J. M. Morales, P. Pinson, H. Madsen, "A bilevel model for electricity retailers' participation in a demand response market environment". *Energy Economics*, vol. 36, pp. 182-197, 2013.

[9] M. Nijhuis, M. Babar, M. Gibescu and S. Cobben, "Demand Response: Social Welfare Maximization in an Unbundled Energy Market Case Study for the Low-Voltage Networks of a Distribution Network Operator in The Netherlands," in IEEE Transactions on Industry Applications, vol. 53, no. 1, pp. 32-38, Jan.-Feb. 2017.

[10] H. Zhong, L. Xie, Q. Xia, "Coupon incentive-based demand response: Theory and case study", *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp.1266-1276, 2012.

[11] M. Song and M. Amelin, "Purchase bidding strategy for a retailer with flexible demands in day-ahead electricity market", *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1839-1850, May 2017.

[12] R. Herranz, A. Munoz San Roque, J. Villar and F. A. Campos, "Optimal demand-side bidding strategies in electricity spot markets", *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1204-1213, Aug. 2012.

[13] C. O. Adika and L. Wang, "Demand-Side Bidding Strategy for Residential Energy Management in a Smart Grid Environment," in IEEE Transactions on Smart Grid, vol. 5, no. 4, pp. 1724-1733, July 2014.

[14] C. Zhang, Q. Wang, J. Wang, P. Pinson, J. M. Morales and J. Østergaard, "Real-time procurement strategies of a proactive distribution company with aggregator-based demand response", *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 766-776, March 2018.

[15] F. Kamyab, M. Amini, S. Sheykhha, et al., "Demand response program in smart grid using supply function bidding mechanism", *IEEE Transactions on Smart Grid*, vol. 7, pp. 1277-84, 2016.

[16] L. Gkatzikis, I. Koutsopoulos and T. Salonidis, "The role of aggregators in smart grid demand response markets", *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1247-1257, July 2013.

[17] V. S. Koritarov, "Real-world market representation with agents", *IEEE Power and Energy Magazine*, vol. 2, no. 4, pp. 39-46, July-Aug. 2004.

[18] Z. Wang and R. Paranjape, "Optimal Residential Demand Response for Multiple Heterogeneous Homes With Real-Time Price Prediction in a Multiagent Framework," in IEEE Transactions on Smart Grid, vol. 8, no. 3, pp. 1173-1184, May 2017.

[19] Z. Zhou, F. Zhao, J. Wang, "Agent-based electricity market simulation with demand response from commercial buildings", *IEEE Transactions on Smart Grid*, vol. 2, pp. 580-8, 2011.

[20] V. Nanduri, T.K. Das, "A reinforcement learning model to assess market power under auction-based energy pricing", *IEEE transactions on Power Systems*, vol. 22, no. 1, pp. 85-95, 2007.

[21] X.Y. Zhang, M. Pipattanasomporn, S. Rahman, "A self-learning algorithm for coordinated control of rooftop units in small- and medium-sized commercial buildings", *Appl. Energy*, vol. 205, pp. 1034-49, 2017.

[22] J. Chen, X. Wang, K. Steemers, "A statistical analysis of a residential energy consumption survey study in Hangzhou, China", *Energy and Buildings*, vol. 66, pp. 193-202, 2013.

[23] M. Kuzlu, M. Pipattanasomporn, S. Rahman, "Impacts of house sizes, appliance ratings and usage patterns on demand response applications: a case-based study". *Intelligent Industrial Systems*, vol. 1, no. 4, pp. 345-57, 2015.

[24] J. Chaney, E. Hugh Owens, A.D. Peacock, "An evidence based approach to determining residential occupancy and its role in demand response management", *Energy and Buildings*, vol. 125, pp. 254-66, 2016.

[25] S. Shao, M. Pipattanasomporn, S. Rahman, "Development of physical-based demand response-enabled residential load models", *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 607-14, 2012.

[26] C.J. Ye, Y. Ding, P. Wang, et al., "A data driven bottom-up approach for spatial and temporal electric load forecasting", *IEEE Transactions on Power Systems*, vol. 3, pp. 1966-1979, 2019.

[27] W. Q. Cui, Y. Ding, H. X. Hui, Z. Z. Lin, P. W. Du, Y. H. Song and C. Z. Shao, "Evaluation and Sequential Dispatch of Operating Reserve Provided by Air Conditioners Considering Lead–Lag Rebound Effect," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6935-50, Nov. 2018.

[28] X. Chen, J. Wang, J. Xie, et al., "Demand response potential evaluation for residential air conditioning loads", *IET Generation, Transmission & Distribution,* vol. 12, pp. 4260-8, 2018.

[29] H. H Hui., Y. Ding, W. Liu, et al., "Operating reserve evaluation of aggregated air conditioners", *Appl. Energy*, vol. 196, pp. 218-28, 2017.

[30] A. Weidlich, D. Veit, "A critical survey of agent-based wholesale electricity market models", *Energy Economics*, vol. 30, pp. 1728-59, 2008.

[31] I. Erev, A. E. Roth, "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria", *American economic review*, vol. 1, pp. 848-81, 1998.

[32] J. Kober, J. A. Bagnell, J. Peters, Reinforcement learning in robotics, A survey. *The International Journal of Robotics Research*, vol. 32, no. 11, 1238-1274, 2013.

[33] M. Zhang, Y. Song, P. Li, et al., "Study on affecting factors of residential energy consumption in urban and rural Jiangsu", *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 330-337, 2016.