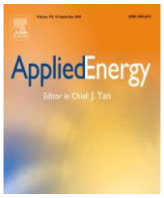
Contents lists available at ScienceDirect

Applied Energy

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



Incremental Incentive Mechanism Design for Diversified Consumers in Demand Response

Di Liu[a], Zhaoming Qin[a], Haochen Hua[b], Yi Ding[c], Junwei Cao[d]

[a] *Department of Automation, Tsinghua University, Beijing, 100084, P. R. China*
[b] *College of Energy and Electrical Engineering, Hohai University, Nanjing, 211100, P. R. China*
[c] *College of Electrical Engineering, Zhejiang University, Hangzhou 310058, China*
[d] *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, 100084, P. R. China*

HIGHLIGHTS:

- An incremental incentive mechanism considering consumer differences is proposed.
- Excessive consumer surplus is avoided through changes of incremental incentive.
- Highly flexible consumers can obtain higher revenue through the redistribution of incentive.
- A model-free approach is proposed to solve the asynchronous optimization problem.

ARTICLEINFO

Keywords:
Incentive-based demand response
Incremental incentive mechanism
Consumer surplus
Deep reinforcement learning

ABSTRACT

Demand response has been proven to be an effective way to improve energy utilization efficiency. It is notable that the diversified characteristics of residential consumers, which many greatly affect its performance in demand response, have not been fully considered in existing incentive mechanisms. In this paper, an incremental incentive mechanism for incentive-based demand response (IBDR) is proposed, in which consumers obtain different incentives according to the increment of response, so that the incentive can follow the change of consumers' marginal cost. We theoretically illustrate that the proposed incremental incentive mechanism can effectively improve the profit of load service entity (LSE), as well as the benefit of highly flexible consumers, compared with other existing incentive mechanism. In practice, LSE's bidding strategy in the day ahead market is affected by the intraday IBDR strategy that cannot be known in advance. In order to solve the bidding problem with incomplete information in the day ahead market, we propose an asynchronous double-interaction deep reinforcement learning (DRL) algorithm to maximize LSE’s cumulative profit of multiple time slots throughout the day. Numerical simulation results show that the proposed mechanism can improve the consumers' response depth while reducing the unit incentive cost, and the proposed DRL algorithm has relatively stable and satisfactory performance even in highly uncertain environment.

1 Introduction

Demand response (DR) has become a critical part of the smart grid, aiming at encouraging consumers to adjust their electricity consumption to improve the efficiency of energy utilization [1]. Generally, large industrial and commercial consumers are considered as better candidates for DR programs, because each consumer can provide considerable response, and the number of consumers is not particularly large, which is conducive to the implementation of DR [2]. However, residential consumers with great demand side flexibility account for approximately 50% of the peak load in many countries, while less than 2% of their flexibility potential is utilized [3]. Since each residential consumer can only provide limited response, it is necessary to

*Corresponding author.

E-mail address: jcao@tsinghua.edu.cn (J. Cao)

Acronyms	
DR	demand response
IBDR	incentive-based demand response
LSE	load service entity
DRL	deep reinforcement learning
WEM	wholesale electricity market
REM	retail electricity market
ISO	independent system operator
IIF	incremental incentive function
MPECs	mathematical programs with equilibrium constraints
TOU	time of use
RTP	real time price
Variables and parameters	
$R_{i,t}$	response of i-th consumer in time slot t
$U_{i,t}$	revenue of i-th consumer in time slot t
$L_{i,t}^{a}$	actual load of the i-th consumer in time slot t
f	the incremental incentive function
g	the unified incentive function
$L_{i,t}^{b}$	base load of i-th consumer in time slot t
$L_{i,t}^{a}$	actual load of i-th consumer in time slot t
$L_{i,t}^{r}$	amount of load rebound of the i-th consumer
$\xi_{i,j}$	load correlation coefficient of the i-th consumer
l_i	comfort loss function of the i-th consumer
$C_{i,t}^{p}$	power purchase cost of the i-th consumer
$C_{i,t}^{c}$	comfort loss
λ_{t}^{TOU}	TOU tariff determined by LSE in advance
$U_{i,t}$	revenue of the i-th consumer from participating in IBDR
L_t	electricity that LSE expects to purchase in time slot t
p_t	purchase price corresponding to L_t
L_{t}^{W}	traded electricity in WEM in time slot t
$C_{W,t}$	cost of LSE in WEM
$C_{S,t}$	penalties paid by LSE
ϕ	penalty function
L_{t}^{S}	electricity made up by ISO
$U_{LSE,t}$	profit of LSE in time slot t
R^{*}	desired response of consumer with incentive function
χ_t	all the parameters of the IIF

aggregate a large number of consumers to form considerable response ability.

In many countries, relatively small consumers cannot directly participate in the wholesale electricity market (WEM) due to the limited ability of WEM to manage a large number of entities. Residential consumers have to participate in the WEM through load service entity (LSE) [4]. Normally, in order to provide energy supply services to consumers in retail electricity market (REM), LSE obtain electricity from the WEM through bidding. The economic benefit of LSE can be improved through DR, including price-based DR and incentive-based demand response (IBDR).

Compared with price-based DR, consumers can directly obtain revenue in incentive-based DR(IBDR) and have higher initiative [5]. Besides, IBDR can provide more flexible dispatchable resources for the power grid, and 93% of the load reduction during the peak period is contributed by IBDR in the US [6]. Therefore, this paper focuses on IBDR issues for residential consumers.

In IBDR, LSE guides consumers to change load consumption through incentive, and consumers participate in response when their cost caused by response can be matched. According to the theory of consumer behavior in microeconomics, the marginal cost of consumers increases with the accumulation of response and varies widely among consumers [2], but in many existing studies on IBDR (e.g.,[7]-[9]), the incentive obtained by consumers are not distinguished according to their differences, which limits the depth of response of highly flexible consumers and reduces the profit that LSE obtain from IBDR.

In order to satisfy consumers' electricity demands, LSE also needs to develop the bidding strategy. In practice, bidding is usually done in the day-ahead market, while IBDR is implemented intraday. Bidding strategy and IBDR strategy influence each other, but in day-ahead market bidding, IBDR strategy is usually not known by LSE in advance, so LSE needs to make bidding decisions in an environment with incomplete information.

Therefore, there are two key issues that need to be addressed. The first is how to make full use of the differentiated characteristics of consumers to improve social benefits on the premise of ensuring fairness. The second is the optimal decision-making method that comprehensively considers day-ahead electricity market bidding and intra-day IBDR.

1.1 Related Works

Surveys and practical experiments in existing studies verify that the response characteristics of consumers are affected by many factors and are quite different from each other [10], [11], which need to be fully considered in IBDR. Some studies attempt to categorize consumers according to their characteristics and then develop incentives separately. Consumer types are considered in [12], and incentives for industrial consumers and residential consumers are formulated separately to elicit different mixtures of IBDR resources with the purpose of minimizing the total procurement cost. Consumers in the same type are further classified into several categories in [13] according to their characteristics, so as to increase their participation in the IBDR and to compensate for the discomfort, similar study has also been reported in [14]. However, the differences of consumers within the category have not been paid special attention.

To solve this problem, an incentive mechanism is proposed in [15], consumers are first set with different weights according to the type, and then the subsidy is calculated according to the contribution of each consumer in IBDR. Similarly, a framework for aggregating residential demands enrolled in IBDR is proposed in [16], in which consumers obtain different incentives based on their actual response. The reward allocation mechanism proposed in [17] can also ensure that consumers obtain different incentives according to their contributions.

Although the above-mentioned IBDR mechanism takes full advantage of the differences of consumers, there are still some shortcomings. When the number of consumers is large, the implementation of IBDR faces a dilemma. Developing differentiated incentive for each consumer is almost infeasible in practice although it can maximize the effect of IBDR, because the detailed characteristic information of each consumer is difficult to be obtained, and the computational efficiency also faces great challenge. Besides, differentiated incentives for consumers may cause issues of fairness.

Besides, the consumers' comfort loss function is nonlinear and concave [10], i.e., comfort loss caused by unit response increases with the response depth. In the existing mechanism, consumers get the same incentive per unit of response, so existing mechanism fails to track changes in consumer comfort loss, resulting in excessive consumer surplus in the early stage of the response [18], thereby reducing the efficiency of IBDR and the profit of LSE.

In addition to the problem of mechanism design, strategy optimization in a coupled environment is also one of the important problems need to be solved. Numerous existing works have been dedicated to solving this problem, mainly converting them to mathematical programs with equilibrium constraints (MPECs) [19][20]. In order to realize the bidding decision with incomplete information, some studies introduce iterative methods [21][22]. However, the diversity of residential consumers makes it difficult to choose an appropriate model and identify corresponding parameters to accurately describe the response behavior of each consumer, which makes traditional solving methods ineffective.

To address the lack of an accurate consumer model in practice, model-free deep reinforcement learning (DRL) is applied in [23] to solve the coupled problem of bidding and IBDR without explicit models, but it is assumed that bidding and IBDR are completed in the same time slot, which is inconsistent with the reality.

1.2 Research gaps and contributions

In summary, the current research has shortcomings in both incentive mechanism and optimization strategy. In terms of incentive mechanism, the contradiction between the full utilization of consumer differentiation and computational efficiency in IBDR has not been well resolved. Since the incremental incentive cost is positively related to the total response of consumers in the existing incentive mechanism, the response depth of highly flexible consumers is limited. Besides, the uneven distribution of consumer surplus also reduces LSE profits. In terms of optimization strategies, since intraday IBDR strategy and day-ahead bidding strategy interact with each other and are executed at different time periods in practice, optimization strategy with incomplete information for day-ahead bidding need to be developed.

The main contribution of this paper is to propose an incentive mechanism to improve the efficiency of IBDR for differentiated residential consumers while ensuring fairness and computational efficiency. The superiority, efficiency and fairness of the proposed incremental incentive mechanism is verified through mathematical analysis and simulation. Meanwhile, an asynchronous double-interaction DRL algorithm is proposed to solve the bidding optimization problem without IBDR information. The main importance and contribution can be summarized as follows.

(1) An incremental incentive mechanism is proposed, in which the incremental unit incentive cost in IBDR is decoupled from the total response. We demonstrate theoretically that the unified incentive mechanism in the existing research is a special form of the proposed incremental incentive mechanism.

(2) Compared with the unified incentive mechanism in existing research [10]-[23], the proposed mechanism can improve the response depth of highly flexible consumers in IBDR, and can balance the consumer surplus per unit of incremental response by tracking the changes in consumer response elasticity to avoid excessive consumer surplus, thus the profit of LSE in IBDR can be improved.

(3) An asynchronous double-interaction DRL algorithm based on deep deterministic policy gradient (DDPG) is proposed to solve the bidding optimization with incomplete information, as well as the difficulty that the parameters of diversified consumer model cannot be accurately identified in practice, so as to maximize the cumulative profit of multiple time slots throughout the day of LSE.

1.3 Organization of the paper

The rest of this paper is organized as follows: Section 2 illustrates the framework of the system, introduces the proposed incremental incentive mechanism and the corresponding models. The superiority, efficiency and fairness of the proposed mechanism are demonstrated in Section 3. Section 4 proposes the asynchronous double-interaction DRL algorithm and numerical simulation is given in Section 5. Finally, the conclusions and further works are drawn in Section 6.

2 System and model

This paper considers LSE that provides energy supply services as well as IBDR program for multiple consumers in REM. Assume one day is decomposed into T time slots and each time slot is represented by t. As shown in Fig. 1, LSE provides power supply services to consumers and bids for the required electricity from WEM. When the clearing price in WEM is high, LSE implements IBDR to improve profit. The power demand of consumers needs to be satisfied in any case, otherwise LSE has to pay penalties due to the power imbalance.

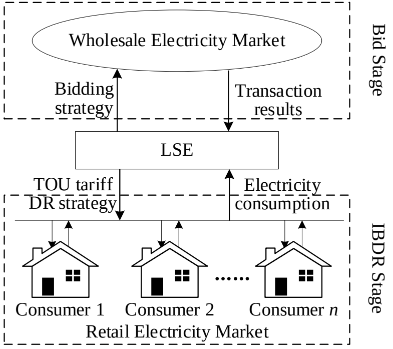


Fig. 1. Bidding and IBDR architecture

2.1 Incremental Incentive Mechanism Formulation

According to the elasticity theory in economics, consumer's price elasticity of response E in IBDR can be defined as follows [24]:

$$E = \frac{dR/R}{dI/I}, \tag{1}$$

where R is the response of consumer, and I is the incentive price. Due to the loss of consumer comfort, E decreases with the increase of R. Meanwhile, due to the diversity of consumers, the changing trend of E among consumers is different. However, in the existing incentive mechanisms, all cumulative response of a consumer in an IBDR event is settled at the same incentive price, in the sense that the changes in E cannot be fully reflected.

In order to trace changes in E, while taking full advantage of the differences among consumers, an incremental incentive mechanism is proposed in this paper, in which consumers obtain revenue based on the increment of response:

$$U_{i,t}(R_{i,t}) = \begin{cases} \int_0^{R_{i,t}} f(x)dx, & R_{i,t} > 0, \\ 0, & others, \end{cases} \tag{2}$$

where f is the incremental incentive function (IIF), and it is monotonically increasing, so that the unit revenue obtained by the consumer increases with the response depth. Note, any function satisfying the above requirements can be used as IIF in the proposed mechanism, e.g., linear function, polynomial function, exponential function, etc.

2.2 Consumer Modeling

Consumers purchase the required electricity from LSE in each time slot according to their power demand and participate in IBDR to improve their benefit. Assuming that the i-th consumer's base load in time slot t is $L_{i,t}^b$, which is also the baseline for calculating the consumer's response in the IBDR. In this sense, the response $R_{i,t}$ of the i-th consumer in time slot t is

$$R_{i,t} = L_{i,t}^b - L_{i,t}^a, \tag{3}$$

where $L_{i,t}^a$ is the actual load of the i-th consumer in time slot t.

When consumers participate in IBDR during time slot t, part of the load is transferred to the subsequent time slot due to the influence of transferable load, e.g., temperature control load, electric vehicle, etc., [25]. Since consumers' load consumption behavior is periodic, their load rebound have similar characteristics in the same time slot [26], and it can be expressed as:

$$L_{i,t}^r = \sum_{j=0}^{t-1} \xi_{i,j} R_{i,j}, \tag{4}$$

where $L_{i,t}^r$ is the amount of load rebound of the i-th consumer, and $\xi_{i,j} \in [0,1]$ is the load correlation coefficient of the i-th consumer, indicating the relationship between the load reduction of the previous time slot and the load rebound of the subsequent time slot. It should be noted that the rebounded load $L_{i,t}^r$ cannot be directly obtained in practice, and only the actual load including the load rebound is required to be known for optimization.

The comfort loss caused by participating in IBDR is related to consumer's response and flexibility, which can be expressed as:

$$C_{i,t}^c = \int_0^{R_{i,t}} l_i(\tilde{R}_{i,t}) d\tilde{R}_{i,t}, \tag{5}$$

where $l_i(R_{i,t})$ is the comfort loss function of the i-th consumer. According to the principles of economics, the marginal cost, i.e., the comfort loss per unit response of consumers is increasing when the load demand is reduced, so $l_i(R_{i,t})$ is a monotonically increasing function, and its form and parameters vary according to the characteristics of consumers.

The power purchase cost of the i-th consumer in time slot t is:

$$C_{i,t}^{p}=\lambda_{t}^{TOU}L_{i,t}^{a}, \tag{6}$$

where λ_t^{TOU} is the TOU tariff determined by LSE in advance.

Consumers obtain the optimal action in each time slot t by solving the following cost minimization problem:

$$\min_{L_{i,t}^{a}}[C_{i,t}^{p}+C_{i,t}^{c}-U_{i,t}], \quad s.t. L_{i,t}^{a}\geq 0, \tag{7}$$

where $U_{i,t}$ is the revenue of the i-th consumer participating in IBDR in each time slot and can be obtained according to (2).

2.3 LSE Modeling

LSE needs to optimize the bidding strategy in WEM and the IBDR strategy in REM. In WEM, LSE buys electricity through bidding, and the bidding strategy $\pi_b(t)$ can be expressed by a monotonically increasing function [27]:

$$p_t=\alpha_t+\beta_t L_t, \tag{8}$$

where L_t is the electricity that LSE expects to purchase at price p_t in time slot t, α_t determines the highest purchase price of LSE, and $\beta_t\leq 0$ determines the trend of the bidding curve. After the WEM is cleared, the electricity whose unit price is higher than or equal to the clearing price of the WEM is traded.

According to (8), the traded electricity L_t^W can be expressed as

$$L_t^W=\left(\lambda_t^W-\alpha_t\right)/\beta_t, \tag{9}$$

where λ_t^W is the clearing price in the WEM, and in the competitive electricity market, its value is not affected by the bidding of a single LSE. The cost of LSE in WEM $C_{W,t}$ in time slot t can be calculated:

$$C_{W,t}=\lambda_t^W\left(\lambda_t^W-\alpha_t\right)/\beta_t. \tag{10}$$

Consumers' power demand should be satisfied in all time slots, and if the traded electricity L_t^W cannot satisfy the consumers' power demand, LSE has to pay penalty to the independent system operator (ISO) who is responsible for compensating the load imbalance. The penalty of LSE related to the electricity shortfall is as follows [23]:

$$C_{S,t}=\phi\left(\sum_{i=1}^{n}L_{i,t}^{a}-L_t^W\right), \tag{11}$$

where $C_{S,t}$ is the penalties paid by LSE, ϕ is the penalty function whose value is positively correlated with the electricity shortfall. Then, the power balance constraint

$$L_t^W+L_t^S=\sum_{i=1}^{M}L_{i,t}^{a}, \tag{12}$$

need to be satisfied, where L_t^S is the electricity made up by ISO after the LSE pays the penalty. Assuming that there are M consumers in the system, the total load of consumers in any time slot should not exceed the maximum capacity L_{max} of the transmission line, i.e.,

$$0\leq\sum_{i=1}^{M}L_{i,t}^{a}\leq L_{max}. \tag{13}$$

The comprehensive profit that LSE obtains from WEM and REM are:

$$U_{LSE,t}=\sum_{i=1}^{n}C_{i,t}^{p}-\sum_{i=1}^{n}U_{i,t}-C_{W,t}-C_{S,t}, \tag{14}$$

where $U_{LSE,t}$ is the profit of LSE in time slot t, $U_{i,t}$ is the revenue obtained by the i-th consumer in time slot t by participating in IBDR.

The optimization goal of LSE is to maximize the cumulative profit of multiple time slots throughout the day, i.e.,

$$\max_{f(R),\alpha,\beta}\sum_{t=0}^{T-1}U_{LSE,t}. \tag{15}$$

In order to ensure that there is no negative incentive in REM, $f(R)$ should be positive and monotonically increasing, and $f(R)\geq 0$, $df(R)/dR\geq 0$ should be satisfied. In WEM, the bidding curve submitted by LSE should be monotonically decreasing, so $\alpha_t\geq 0$, $\beta_t\leq 0$ should be satisfied.

3 Analysis of Incremental Incentive Mechanism

In this section, we theoretically analyze the superiority of the proposed incremental incentive mechanism, and introduce the Stackelberg game theory to analyze the fairness and Pareto efficiency of the proposed mechanism.

3.1 Superiority Analysis

In this part, we theoretically illustrate the advantages of the proposed incremental incentive mechanism through two propositions.

We first show that the existing unified incentive mechanism is a special form of incremental incentive mechanism. When IIF is independent of response increment, i.e., $f(R)$ is a constant, let $f(R)=I$, the revenue obtained by consumers is

$$\int_0^{R_{i,t}}f(x)dx=\int_0^{R_{i,t}}Idx=IR_{i,t}, \tag{16}$$

where I is a constant and represents the incentive in the IBDR. Therefore, the consumers' revenue is the product of unit incentive I and total response $R_{i,t}$, which is the same as unified incentive mechanism.

The following two propositions discuss the advantages of the proposed incremental incentive mechanism from two aspects: changes in consumer surplus of individual consumer during the accumulation of response, and the distribution of response and consumer surplus among consumers.

Proposition 1: In an IBDR event of a time slot, let $R^*_{1,f}$ and $R^*_{1,g}$ be the optimal response of consumer 1 in the proposed incremental incentive mechanism f and existing unified incentive mechanism g, respectively. LSE needs to pay the corresponding incentive according to the consumer's response, denoted as $C_f\left(R^*_{1,f}\right)$ and $C_g\left(R^*_{1,g}\right)$ under the two mechanisms, respectively. Then we have

$$C_f(R^*_{1,f}) < C_g(R^*_{1,g}), \qquad \forall R^*_{1,f} = R^*_{1,g}. \tag{17}$$

Proof: Assuming that $f(R)$ is an IIF, which is an increment function, $g(R)$ is the unified incentive function, which remains constant in a certain IBDR event. According to the existing research, the comfort loss of consumers caused by per unit load reduction is increasing [12]. Similar to the supply curve in economics, the consumer's load reduction function can be expressed as $I = z(R)$, which represents the relationship between the unit incentive and the desired load reduction. In IBDR, consumers obtain revenue according to the incentive function $f(R)$. Assuming that consumers are rational, their decision on load reduction is to maximize the difference between economic benefits and comfort loss, as follows:

$$R^* = arg\,\max_R \int_0^R [f(x) - z(x)]dx \tag{18}$$

where R^* is the desired response of consumer with the incentive function $f(R)$, $\int_0^R [f(x) - z(x)]dx$ is the consumer surplus with response R [18]. When the consumer surplus is positive, consumers can obtain benefit through response and choose to participate in the IBDR. Since the comfort loss of consumer is positively related to the response, $z(R)$ is a monotonically increasing function. Considering that there are huge differences in the characteristics of consumers, and their response behaviors are affected by many factors, $z(R)$ may have many forms.

As shown in Fig. 2, suppose that in a specific IBDR event, $f(R)$, $g(R)$ and $I_1(R)$ intersect at the same point. In this sense, the response of consumers is the same in the two incentive mechanisms, i.e., $R^*_{1,f} = R^*_{1,g}$.

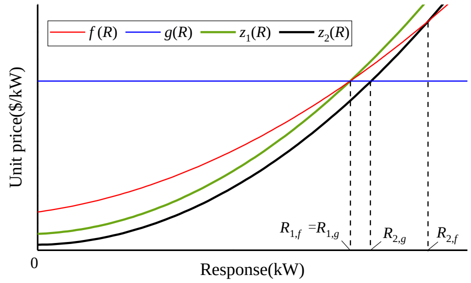


Fig. 2. Schematic diagram to illustrate the incremental incentive mechanism

The cost of LSE in the two incentive mechanisms are

$$C_f(R^*_{1,f}) = \int_0^{R^*_{1,f}} f(x)dx, \tag{19a}$$

$$C_g(R^*_{1,g}) = g(R^*_{1,g})R^*_{1,g}. \tag{19b}$$

By deriving (19), the incremental costs of the two incentive mechanisms can be obtained as

$$\Delta C_f(R) = \frac{dC_f(R)}{dR}\Delta R = f(R)\Delta R, \tag{20a}$$

$$\Delta C_g(R) = \frac{dC_g(R)}{dR} = g(R)\Delta R, \tag{20b}$$

where ΔR is the increment of response. Since $f(R)$ is a monotonically increasing function, the following inequality always holds:

$$f(R) < g(R), \;\; \forall R < R^*_{1,f}. \tag{21}$$

With (18)-(21), it is easy to show that (17) holds. ■

Proposition 1 describes the relationship between LSE and consumers in IBDR. It shows that the cost in incremental incentive mechanism is always lower than that in the existing unified incentive mechanism when consumer has the same response. Incremental incentive mechanism can trace the changes in consumer response flexibility better, so that the consumer surplus obtained by the consumer per unit of response is maintained in a reasonable range in the process of response accumulation, thereby reducing the incentive cost.

Proposition 2: Assuming that there are two consumers with different flexibility, the response function of low-flexible consumer and high-flexible consumer are represented by $I_1 = z_1(R)$ and $I_2 = z_2(R)$, respectively. Let $R^*_{2,f}$ and $R^*_{2,g}$ denote the response of high-flexible consumer with incentive function $f(R)$ and $g(R)$, respectively. The following two formulas always hold:

$$R^*_{2,f}/R^*_{1,f} > R^*_{2,g}/R^*_{1,g}\,, \forall R^*_{1,f} = R^*_{1,g}, \tag{22}$$

$$\frac{\int_0^{R^*_{2,f}}[f(x) - z(x)]dx}{\int_0^{R^*_{1,f}}[f(x) - z(x)]dx} > \frac{\int_0^{R^*_{2,g}}[g(x) - z(x)]dx}{\int_0^{R^*_{1,g}}[g(x) - z(x)]dx}, \forall R^*_{1,f} = R^*_{1,g}. \tag{23}$$

Proof: As shown in Fig. 2, the low-flexible consumer performs the same response in the two mechanisms. According to (18), we can obtain $R^*_{2,g}$ by solving $g(R) = z_2(R)$. Since $f(R)$ is a monotonically increasing function and $f\left(R^*_{1,f}\right) = g\left(R^*_{1,g}\right)$, we have

$$f(R) > g(R), \qquad \forall R > R^*_{1,g}. \tag{24}$$

By substituting $R^*_{2,g}$ and (24) into (18), we can obtain the change in consumer surplus generated by the incremental response:

$$\left.\frac{d\int_0^R [f(x) - z(x)]dx}{dR}\right|_{R=R^*_{2,g}} = f(R^*_{2,g}) - z(R^*_{2,g}) > 0. \tag{25}$$

Since increasing the response at $R^*_{2,g}$ increases benefit, the high-flexible consumer will increase the response until the consumer surplus increment is zero, i.e., $f(R) = z(R)$. Therefore, $R^*_{2,f}$ is obtained by solving $f(R) = z(R)$ and is larger than $R^*_{2,g}$, thus (22) holds. Besides, according to (22) and (24), we have

$$\begin{cases} \int_{R_{1,f}^*}^{R_{2,f}^*}[f(x)-z(x)]dx > \int_{R_{1,g}^*}^{R_{2,g}^*}[f(x)-z(x)]dx, \\ \int_{0}^{R_{1,f}^*}[f(x)-z(x)]dx < \int_{0}^{R_{1,g}^*}[g(x)-z(x)]dx, \end{cases} \quad (26)$$

and (23) holds. ■

Proposition 2 describes the relationship among consumers in IBDR. It illustrates the increment incentive mechanism can increase the response depth of high-flexible consumers and transfer more consumer surplus to them. With the increment incentive mechanism, consumers who actively participate in load reduction can get a higher percentage of benefit, thereby enhancing consumers' enthusiasm for participating in IBDR.

3.2 Efficiency and Fairness Analysis

Fairness is the basic requirement of an IBDR mechanism. According to [28], the most important fairness axioms are (i) sharing incentive, (ii) Pareto efficiency, (iii) strategy-proofness and (iv) envy-freeness. The fairness of the proposed mechanism is analyzed as follows:

(i) Sharing Incentive: In IBDR, sharing incentive means that the revenue is allocated to different consumers based on their response, rather than simply distributing the revenue equally. With sharing incentive, if the response of consumer i is larger than that of consumer j in time slot t, then consumer i must obtain higher revenue. In (2), since $f(\tilde{R}_{i,t})$ is a monotonically increasing function, the consumer's revenue increases monotonically with respect to the response R, i.e., if $R_{i,t} > R_{j,t}$, then $\int_0^{R_{i,t}} f(x)dx > \int_0^{R_{j,t}} f(x)dx$. Thereby, $U_{i,t} > U_{j,t}$ always holds.

(ii) Pareto Efficiency: The Stackelberg game model can be used to describe the interactive relationship between LSE and consumers, where LSE is the leader and formulates incentive strategies, and the consumer is the follower, making response decisions based on incentives in different time slots. When the Stackelberg game meets the following conditions, there exists a unique equilibrium solution [12], [29]. (a) The revenue of consumers has a unique maximum once informed of the strategy of LSE. (b) The profit of LSE has a unique maximum for a given strategy of the consumers. According to the principle of backward induction, the Pareto efficiency of the proposed mechanism is analyzed as follows [12]:

We first analyze the optimal decision-making process of consumers in (a). As discussed above, consumers participate in IBDR only when the revenue gained by consumers are larger than their loss of comfort. The benefits obtained by consumers participating in IBDR are:

$$U_{i,t}(R_{i,t}) = \int_0^{R_{i,t}} [f_t(x) - z_{i,t}(x)]dx. \quad (27)$$

The first derivative of $U_{i,t}(R_{i,t})$ with respect to $R_{i,t}$ is

$$\frac{dU_i}{dR_{i,t}} = f(R_{i,t}) - z_i(R_{i,t}), \quad (28)$$

where, $f(R_{i,t})$ and $z_i(R_{i,t})$ are both continuous and differentiable increasing functions.

If $f(0) > z_i(0)$ and there exist $R_{i,c,t} \in [0, R_{i,max}]$, let $f(R_{i,c,t}) = z_i(R_{i,c,t})$, then we have

$$\begin{cases} f(R_{i,t}) - z_i(R_{i,t}) \geq 0, \forall R_{i,t} \in [0, R_{i,c,t}], \\ f(R_{i,t}) - z_i(R_{i,t}) \leq 0, \forall R_{i,t} \in [R_{i,c,t}, R_{i,max}]. \end{cases} \quad (29)$$

The second derivative of $U_{i,t}$ with respect to $R_{i,t}$ at $R_{i,c,t}$ is

$$\frac{d^2U_i}{dR_i^2} = f'(R_i) - z_i'(R_i), \quad (30)$$

According to the definition of derivative, (30) can be rewritten as

$$\frac{d^2U_i}{dR_i^2} = \frac{f(R_{i,c,t}) - z_i(R_{i,c,t})}{\Delta R} - \frac{f(R_{i,c,t} - \Delta R) - z_i(R_{i,c,t} - \Delta R)}{\Delta R}, \quad (31)$$

According to (31), $d^2U_i/dR_i^2 < 0$ holds, and $U_{i,t}$ reaches the maximum value at $R_{i,c,t}$, i.e., $R_{i,t}^* = R_{i,c,t}$.

If $f(0) > z_i(0)$ and $R_{i,c,t} \notin [0, R_{i,max}]$, let $f(R_{i,c,t}) = z_i(R_{i,c,t})$, consumer i participating in IBDR can always obtain positive incremental benefit, so they participate in IBDR as much as possible, i.e., $R_{i,t}^* = R_{i,t,max}$.

If $f(0) \leq z(0)$, consumer i do not participate in IBDR, i.e., $R_{i,t}^* = 0$.

Next, we analyze the decision-making behavior of LSE in (b). Let χ_t denote the incremental incentives, and the first derivative of (14) with respect to χ_t is:

$$\frac{\partial U_{LSE}}{\partial \chi_t} = \sum_{i=1}^{n} \left(\left(\lambda_t^W - \lambda_t^{TOU} - f(R_{\mathrm{i,t}}^*) \right) \frac{\partial R_{\mathrm{i,t}}^*}{\partial \chi_t} \right). \quad (32)$$

The consumer's response $R_{i,t}^*$ increases with the unit incentive price obtained, i.e., $\partial R_{i,t}^*/\partial \chi_t \geq 0$.

If $\lambda_t^W \leq \lambda_t^{TOU}$, $U_{LSE,t}$ decreases monotonically with χ_t, so LSE does not perform IBDR.

If $\lambda_t^W > \lambda_t^{TOU}$, we have

$$\begin{cases} \dfrac{\partial U_{LSE}}{\partial \chi_t} > 0, \lambda_t^W > \lambda_t^{TOU} + f(R_{\mathrm{i,t}}^*), \\ \dfrac{\partial U_{LSE}}{\partial \chi_t} \leq 0, \lambda_t^W \leq \lambda_t^{TOU} + f(R_{\mathrm{i,t}}^*), \end{cases} \quad (33)$$

$U_{LSE,t}$ achieves the maximum value at $\lambda_t^W = \lambda_t^{TOU} + f(R_{i,t}^*)$.

In summary, there exists a unique equilibrium in the proposed mechanism, which achieves Pareto optimization.

(iii) Strategy-Proofness: In a strategy-proof mechanism, consumers cannot obtain additional benefits from reporting false information. Since the only information that consumers need to report is their response, any false response will deviate from its Pareto optimal value, i.e.,

$$\left(C_{i,t}^{p}+C_{i,t}^{c}-U_{i,t}\right)\Big|_{R_{i,t}=R_{\mathrm{i,t}}^{*}} < \left(C_{i,t}^{p}+C_{i,t}^{c}-U_{i,t}\right)\Big|_{R_{i,t}=R_{\mathrm{i,t}}^{-}}, \quad (34)$$

where $R_{\mathrm{i,t}}^{-}$ refers to any possible value except $R_{\mathrm{i,t}}^{*}$. Hence, the mechanism is strategy-proofness.

(iv) Envy-Freeness: In an envy-free mechanism, no consumer envies another consumer's allocation. Specifically, consumers get the same revenue under the same response. From (11), the following equation always holds:

$$\int_{0}^{R_{i,t}} f(x)dx = \int_{0}^{R_{j,t}} f(x)dx\,, \forall R_{i,t} = R_{j,t}, \qquad (35)$$

hence, the designed mechanism is envy-free.

4 Problem Formulation and Solution

In this section, the decision sequence of bidding and IBDR is first clarified. Then, the coupled problem in WEM and REM is formulated into an MDP. Finally, an asynchronous double-interaction DRL algorithm is proposed to solve the optimization problem.

4.1 Decision Timeline and Methodology

Although bidding and IBDR decisions are coupled with each other, in practice, they are executed in different time slots. The bidding decision is made in the day-ahead market, while the IBDR is executed intraday, as shown in Fig. 3.

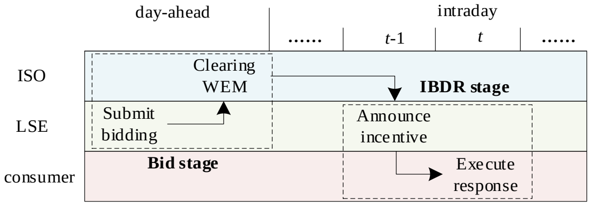


Fig. 3. Timelines of actions

In a day-ahead market, LSE submits bid for each time slot of the next day to ISO, and then ISO clears the WEM based on the bids/offers of the buyer and the seller to yield the wholesale electricity price, as well as the electricity transacted by each participant. Prior to time slot t intraday, LSE announces the IBDR incentive to the consumers for the next time slot, and consumers execute the response accordingly.

Theoretically, this couped problem can be transformed into a two-layer optimization problem, where the upper layer is a bidding problem, and the cleared price and electricity are used as input to the IBDR problem in the lower layer. Besides, the optimization result of the lower layer also affects the decision-making of the upper layer. However, the parameters of the comfort loss function of each consumer are difficult to be obtained, so the optimization decision cannot be directly calculated.

DRL and heuristic algorithms, such as genetic algorithm, particle swarm algorithm, etc., can find the optimal solution in a model-free environment. As shown in Fig. 3, bidding and IBDR are asynchronous decision-making processes. For online algorithms, solving problems requires complete information. However, the actions and optimization results of IBDR are unknown when solving the bidding problem. Therefore, online heuristic algorithms are difficult to solve the coupling problem in this paper.

The training process of the DRL algorithm is offline, and the network can be trained with complete historical information, after the bidding and IBDR are both completed. Besides, benefiting from the learning and memory ability of the neural network, DRL can infer the possible optimization results of IBDR with incomplete information and make the optimal bidding decision, accordingly. Considering that the action space of bidding and IBDR is continuous, it is necessary to apply a model-free policy-based DRL algorithm to solve the coupled problem. It has been reported that deep deterministic policy gradient (DDPG) algorithm has relatively good performance on the prediction accuracy and convergence speed among the model-free policy-based DRL algorithm, but it requires more state transition samples [30], [31]. Since electricity consumption data can be easily collected by smart meters, in the paper, we propose the asynchronous double-interaction DRL algorithm based on the DDPG to solve the coupled problem.

4.2 Markov Decision Process Formulation

The coupled problem can be formulated as a MDP which is formally defined as a five-tuple $M = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{T}$ is the transition probability between states, $\mathcal{R}$ is the reward function, and γ is the discount factor.

Since WEM is the day-ahead market, the information related to the WEM available to the LSE includes predicted price $\hat{\lambda}_t^{W}$ in WEM, TOU price λ^{TOU}, and predicted load demand without IBDR $\hat{L}_\tau^{o}$ for the next day. Let $s_{W,t}$ denote the set of information available to the LSE before executing biding action $a_{W,t}$ in WEM, which can be expressed as:

$$s_{W,t} = \left(\hat{\lambda}_\tau^{W}, \lambda_\tau^{TOU}, \hat{L}_\tau^{o}\right), \quad \tau = 1,2,\ldots,T. \qquad (36)$$

IBDR is implemented intraday, and the information related to the REM available to the LSE includes the actual load and price in WEM before the current time slot, the predicted load demand $\hat{L}_t$, the clearing price λ_τ^{W} in WEM, and the TOU price λ_t^{TOU}. Let $s_{R,t}$ denote the set of information available to the LSE before executing IBDR action $a_{R,t}$ in REM, which can be expressed as:

$$s_{R,t} = \left(\lambda_\tau^{W}, \lambda_\tau^{TOU}, L_\kappa^{a}, \hat{L}_\tau^{o}\right), \tau = 1,\ldots,T, \kappa = 1,\ldots,t-1. \quad (37)$$

In the considered scenario, LSE needs to decide the bidding strategy in WEM, that is, to decide the values of α_t and β_t. Meanwhile, the IIF $f(R)$ also needs to be determined. Due to the different forms of IIFs, the number of parameters that need to be decided is different. Let χ_t represent all the parameters of the IIF, and the action a_t is defined as

$$a_t = (\alpha_t, \beta_t, \chi_t). \tag{38}$$

Since our goal is to increase the profit of LSE, let us define the reward r_t in time slot t as the profit obtained by the LSE:

$$r_t = \sum_{i=1}^{n} C_{i,t}^{p} - \sum_{i=1}^{n} U_{i,t} - C_{w,t} - C_{S,t}. \tag{39}$$

Inspired by the credit allocation mechanism [32], let us redistribute the reward of bids and IBDR, instead of directly using the overall reward r_t. Since the cost of IBDR does not affect bidding, the reward for bidding $r_{W,t}$ is defined as:

$$r_{W,t} = \sum_{i=1}^{n} C_{i,t}^{p} - C_{w,t} - C_{S,t}. \tag{40}$$

Similarly, since the penalty caused by load imbalance has nothing to do with IBDR, the reward of IBDR $r_{R,t}$ is defined as:

$$r_{R,t} = \sum_{i=1}^{n} C_{i,t}^{p} - \sum_{i=1}^{n} U_{i,t} - C_{w,t}. \tag{41}$$

The cumulative discounted reward from time slot t and onwards is donated by D_t and can be express as

$$D_t = \sum_{k=t}^{T-1} \gamma^{k-t} r_t, \tag{42}$$

where $\gamma \in (0,1]$ is the discount factor.

4.3 Asynchronous Double-Interaction DRL Algorithm

LSE needs to sequentially make the optimal decision in WEM and REM, i.e., the bidding strategy $\pi_W(s_{W,t}) = (\alpha_t, \beta_t)$ with the observation s_t^{BID}, and the IBDR strategy $\pi_R(s_{R,t}) = \chi_t$ with the observation $s_{R,t}$. Although bidding and IBDR are executed sequentially, the two issues are coupled. The formulation of bidding strategies requires the optimization results of IBDR, which cannot be known in advance. Besides, the reward is calculated jointly by the bidding and IBDR, i.e., the reward cannot be immediately obtained when the bidding action is completed. In order to solve the above problems, we propose an asynchronous double interaction DRL algorithm, in which the network is trained offline after the bidding and IBDR actions are both completed. The interaction between each component in the algorithm is illustrated in Fig. 4.

Let G_W and G_R represent sub agents in WEM and REM, respectively. On the previous day, G_W first performs action $a_{W,t}$ in WEM according to $\pi_W(s_{w,t})$. Then, in the time slot $t-1$, G_R performs action $a_{R,t}$ in REM according to $\pi_R(s_{R,t})$, and the reward value r_t can be calculated accordingly.

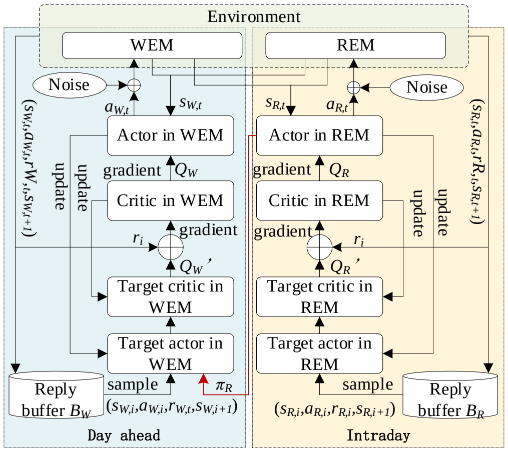


Fig. 4. Asynchronous double-interaction DRL algorithm

After the bidding and IBDR are completed, the two sets of state transitions in WEM and REM, $(s_{W,t}, a_{W,t}, r_{W,t}, s_{W,t+1})$ and $(s_{R,t}, a_{R,t}, r_{R,t}, s_{R,t+1})$, are stored in the replay buffer B_W and B_R, respectively. Each replay buffer can store a population of K samples, and when it is full, the oldest sample would be eliminated. Since the data used to train the network needs to be independent and identically distributed, the data is randomly sampled from the replay buffer to train the network. In the proposed DRL algorithm, the interaction between the two sub-agents and the environment is coupled, but the network training is independent. Since the real load is affected by IBDR, it is necessary to train the IBDR agent first, and then use the IBDR agent as the environment to participate in the training of the bidding agent. According to the optimization strategy π_R in REM, the real load demand after IBDR can be predicted:

$$\hat{L}_t^{o} \xrightarrow{\pi_R} \hat{L}_t^{a}, \tag{43}$$

i.e., with the IBDR action strategy π_R, the real load of consumers after participating in IBDR can be predicted.

LSE desires to maximize the total profit for the entire period, and the goal of the proposed DRL algorithm is to learn a policy π to maximize the expected return of actions in the initial slot, i.e., $J[\pi] = \mathbb{E}_{a_t \sim \pi}[D_0]$, where $J[\pi]$ is the objective function and $\mathbb{E}$ is mathematical expectation, D_0 is cumulative return from time slot 0.

In policy-based DRL algorithm, the action value function is used to measure the performance of policy π,

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{a_{i>t} \sim \pi}[D_t | s_t, a_t], \tag{44}$$

where $Q^{\pi}(s_t, a_t)$ represents the total expected return of each action after time slot t. Calculate the return of each step is obviously time consuming. To improve the training efficiency of the algorithm, (44) can be replaced by a recursive form according to the Bellman equation

$$Q(s_t, a_t) = \mathbb{E}\left[r_t + \gamma Q\left(s_{t+1}, \mu(s_{t+1})\right)\right], \tag{45}$$

where $\mu(s_{t+1})$ is the actor function which specifies the current policy by mapping states to a specific action.

Let θ^{Q} and θ^{μ} denote the parameter vectors of critic network and action network, respectively. Besides, let $\theta^{Q'}$ and $\theta^{\mu'}$ denote the parameter vectors of target critic network and target action network, respectively. The parameter vectors of critic network θ^{Q} is updated by minimizing the following loss function:

$$L(\theta^{Q})=\frac{1}{Z}\sum_{i}\left(y_i-Q(s_i,a_i|\theta^{Q})\right)^2, \tag{46}$$

where Z is the number of samples taken from the replay buffer, and $y_i=r_i+\gamma Q'(s_{i+1},\mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'})$. The parameter vectors of actor network θ^{Q} is updated using sampled policy gradient:

$$\nabla_{\theta^{\mu}}J\approx\frac{1}{Z}\sum_{i}\nabla_{a}Q\left(s,a;\theta^{Q}\Big|_{s=s_i,a=\mu(s_i)}\nabla_{\theta^{\mu}}\mu\left(s;\theta^{\mu}\big|_{s_i}\right)\right), \tag{47}$$

where $\nabla_{\theta^{\mu}}$ is the gradient of θ^{μ}, and ∇_{a} is the gradient of a.

In order to improve the exploration efficiency of action space, the exploration policy μ' is constructed by adding noise sampled from a noise process $\mathcal{N}$ to the actor policy

$$\mu^{'}(s_t)=\mu(s_t|\theta_t^{\mu})+\mathcal{N}_t, \tag{48}$$

where Ornstein-Uhlenbeck process $\mathcal{N}_{\mathrm{t}}$ is employed for the noise process [33]:

$$\mathcal{N}_{t+1}=(1-\theta)\mathcal{N}_t+\sigma dW_t\mathcal{N}_t, \tag{49}$$

where $\theta>0$ and $\sigma>0$ are parameters in drift and diffusion terms, respectively, and W_t denotes the Wiener process.

The detailed asynchronous double-interaction DRL algorithm is presented in Algorithm 1.

Algorithm 1 Asynchronous double-interaction DRL algorithm

```
Randomly initialize parameter vectors $\theta_W^Q$, $\theta_W^\mu$, $\theta_R^Q$, $\theta_R^\mu$.
Initialize target network ${\theta_W}^{Q'} \leftarrow \theta_W^Q$, $\theta_W^{\mu'} \leftarrow \theta_W^\mu$, ${\theta_W}^{Q'} \leftarrow \theta_W^Q$, $\theta_R^{\mu'} \leftarrow \theta_R^\mu$.
Initialize replay buffer $B_W$ and $B_R$.
For episode=1,..,$M$ do
  Initialize random process $\mathcal{N}_W$ for action $a_W$ exploration and $\mathcal{N}_R$ for $a_R$
  Receive initial observation state $s_{W,t}$ and $s_{R,t}$
  For $t=0,\dots,T-1$ do
    Select $a_{W,t}=\mu(s_{W,t}|\theta_W^\mu)+\mathcal{N}_{W,t}$
    Execute $a_{W,t}$ in WEM and receive $L_{i,t}^b$
    Execute $a_{R,t}$ in REM and receive reward $r_t$ and $s_{t+1}$
    Store transition $(s_{W,t},a_{W,t},r_t,s_{W,t+1})$ in replay buffer $B_W$ and
    $(s_{R,t},a_{R,t},r_t,s_{R,t+1})$ in $B_R$
    If the IBDR training is not completed
      Sample a random minibatch of $Z_R$ transitions from replay buffer $B_R$
      Set $y_i=r_i+\gamma_R Q_R^{'}\left(s_{R,i+1},\mu_R^{'}\left(s_{R,i+1};\theta_R^{\mu'}\right);\theta_R^{Q'}\right)$
      Update the critic network of $G_W$ and $G_R$ respectively by minimizing
      $L$ in (46):
          $\theta_R^Q=\theta_R^Q-\nabla_{\theta_R^Q}L(\theta_R^Q)$
      Update actor network of $G_W$ and $G_R$ respectively using sampled
      gradients in (47):
          $\theta_R^\mu=\theta_R^\mu+\nabla_{\theta_R^Q}J(\theta_R^Q)$
      Update the target networks of $G_W$ and $G_R$ respectively:
          $\theta_R^{Q'}\leftarrow\tau_R\theta_R^Q+(1-\tau_R)\theta_R^{Q'}$, $\theta_R^{\mu'}\leftarrow\tau_R\theta_R^\mu+(1-\tau_R)\theta_R^{\mu'}$
    Else
      Sample a random minibatch of $Z_W$ transitions from replay buffer
      $B_W$
      Obtain the IBDR result according to $\pi_R$ and add it to observation
      $s_{W,t}$: $\hat{L}_t^o\rightarrow\hat{L}_t^a$
      Set $y_i=r_i+\gamma_W Q_W^{'}\left(s_{W,i+1},\mu_W^{'}\left(s_{W,i+1};\theta_W^{\mu'}\right);\theta_W^{Q'}\right)$
      Update the critic network of $G_W$ and $G_R$ respectively by minimizing
      $L$ in (46):
          $\theta_W^Q=\theta_W^Q-\nabla_{\theta_W^Q}L(\theta_W^Q)$
      Update actor network of $G_W$ and $G_R$ respectively using sampled
      gradients in (47):
          $\theta_W^\mu=\theta_W^\mu+\nabla_{\theta_W^Q}J(\theta_W^Q)$
      Update the target networks of $G_W$ and $G_R$ respectively:
          $\theta_W^{Q'}\leftarrow\tau_W\theta_W^Q+(1-\tau_W)\theta_W^{Q'}$, $\theta_W^{\mu'}\leftarrow\tau_W\theta_W^\mu+(1-\tau_W)\theta_W^{\mu'}$
  End for
End for
```

5 Numerical Simulation

5.1 Simulation Setup

The RTP λ_t in WEM is taken from PJM electricity market. The number of consumers is set to be 25, and the real load data in Pecan Street [34] is used as the baseline load of consumers, where 92 days are selected as the training set, 10 days as the validation set, and 3 days as the test set. The consumer response function is set as $z_i(R)=a_iR^2+b_iR+c_i$, where a_i, b_i and c_i are randomly generated with uniform distribution within the range [0,0.3], [0,0.1] and [0,0.02], respectively. We set up the IIF $f(R)=\alpha_1+\beta R$ and the unified price $f(R)=\alpha_2$. The rebound coefficient $\xi_{i,t}$ is randomly generated with uniform distribution within the range [0,1], and the penalty function is set to be two times of clearing price in WEM, i.e., $\phi=2\lambda_t^{DA}L_{t,s}$, where $L_{t,s}$ is the electricity shortage in time slot t. In order to simulate the uncertainty caused by the prediction error, a disturbance is added to the real value to simulate the predicted value, i.e., $\hat{\lambda}_t^{DA}=(1+\varepsilon)\lambda_t^{DA}$, and $\hat{L}_t^a=(1+\varepsilon)L_t^a$, where ε is the error generated by the normal distribution with mean and variance in TABLE I. The TOU tariff for each slot of the day is shown in Table II.

TABLE I MEAN AND VARIANCE

Low Uncertainty Scenario		High Uncertainty Scenario	
Mean	Variance	Mean	Variance
0	0.05	0	0.25

TABLE II TOU TARIFF FOR EACH TIME SLOT

0.03$/kWh (Peak)	0.025$/kWh (Flat)		0.02$/kWh (Valley)	
11:00-19:00	6:00-10:00	20:00-22:00	0:00-5:00	23:00

All networks in the DRL adopt the fully connected network with 3 hidden layers, of which the first two have 256 neurons, and the third has 128 neurons. The learning rate of actor network and critic network for both sub-agents are set to be 0.000001 and 0.00001. The algorithm is implemented using *PyTorch* in

Python. The case studies have been performed on a laptop with Intel(R) Core(TM) i7-9750H processor and one single NVIDIA GeForce GTX 1660 Ti GPU.

5.2 Performance Analysis of the Proposed Mechanism

In this part, we test the performance of the proposed incremental incentive mechanism. Since the RTP of the electricity market has large fluctuations and is quite different among days, we chose three consecutive days for simulation.

In REM, LSE provides energy supply services to consumers with fixed TOU tariff. In most time slots, the TOU price in REM is higher than the RTP in WEM, so LSE can obtain profit through power trading. Therefore, in order to improve the total profit, LSE implements IBDR according to the price fluctuation in WEM. The response under different mechanisms during each slot is shown in Fig. 5.

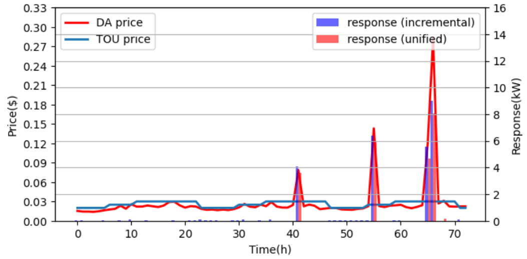


Fig. 5. Response under different mechanisms in each time slot

It can be seen from Fig. 5 that the algorithm proposed in this paper can trace the fluctuation of RTP in WEM. In most time slots, TOU price is higher than RTP, and IBDR is not implemented. In some time slots with high RTP, e.g., time slot 41, 55, 65, 66, etc., IBDR is implemented to reduce the deficits. Since the incentive cost of IBDR is increasing, the depth of IBDR in different time slots also changes with RTP. For example, the RTP in time slot 66 is as high as 0.29$/kW, while the TOU price is only 0.03$/kW, which means that LSE losses 0.26$/kW of electricity providing to consumers. LSE needs to reduce consumers' power consumption as much as possible through IBDR, and the consumers' response exceeds 8kW. In comparison, the RTP in time slot 41 is 0.08$/kW, 0.05$/kW higher than TOU price, with the response less than 4kW.

Further analysis of the unit cost and response of IBDR with different mechanisms are shown in Fig. 6.

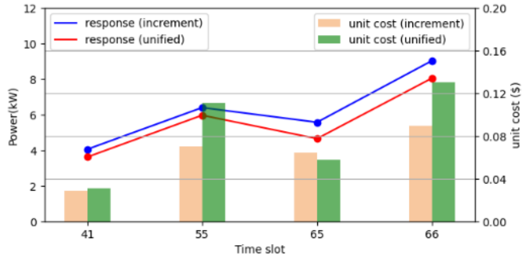
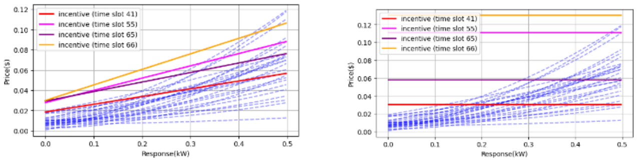


Fig. 6. Unit cost and response with different mechanisms

It can be seen from Fig. 6 that in the four typical time slots, the consumer response with IIF is always higher than that with the unified incentive price. Meanwhile, in time slots 41, 55, and 66, the unit incentive cost with IIF is also lower than the cost with the unified incentive price, indicating that IIF can effectively reduce the incentive cost of LSE and fully release the potential of load reduction of consumers at the same time. In time slot 65, the cost with IIF is slightly higher than that with the unified incentive price, which is caused by the large gap in consumer response. The difference of consumer response in time slot 65 and time slot 66 is almost the same, but there is a significant difference in unit cost, indicating that the difference between unit cost with IIF and unified incentive price increases with the raise of consumer response. This phenomenon is caused by the consumer's response characteristics.



(a) IIF (b) Unified incentive price

Fig. 7. Incentive and consumer response curve

In Fig.7, each blue dotted line represents a consumer's load reduction function $z(R)$, which represents the relationship between the unit incentive and the desired load reduction. The solid lines represent the incentive functions in time slots 41, 55, 65, and 66, respectively. The difference between the incentive function and the consumer response function is consumer surplus.

It can be seen from Fig. 7 (b) that the unified incentive price causes a large consumer surplus at the initial stage of load reduction, and it decreases rapidly with the increase of response. When LSE hopes to obtain higher load reduction, it can only shift the horizontal incentive price curve upward, which exacerbates the problem of uneven distribution of consumer surplus. For example, the consumer surplus in time slot 65 is about 0.04$/kW at the initial stage of load reduction, while it is as high as 0.1$/kW in time slot 66.

The IIF proposed in this paper can effectively alleviate the above problem. With IIF, the largest consumer surplus in time slot 65 is only 0.01$/kW, and in time slot 66 is only about 0.02$/kW. The redistribution of consumer surplus improves the incentive efficiency of LSE, enabling it to obtain more load reduction with lower unit incentive costs, as shown in Fig. 7.

5.3 Performance Analysis of the Proposed Algorithm

Since LSE needs to pay penalty for load imbalance, the bidding in WEM largely affects the profit of LSE. The power obtained from WEM by LSE and the actual electricity consumption demand of consumers in each time slot are shown in Fig. 8.

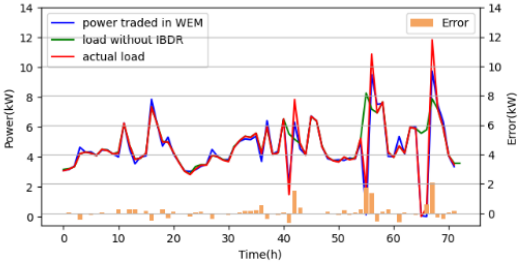


Fig. 8. Traded electricity in WEM and power consumption of consumers

As can be seen from Fig. 8, the proposed asynchronous double-interaction DRL algorithm can ensure that the traded electricity is close to the actual power demand of consumers, even in the time slot with the influence of IBDR and load rebound. For example, in time slot 55, the traded electricity decreases with the actual load demand of consumers after IBDR, while in the subsequent time slot 56, the traded electricity increases with the load rebound. In the optimization of the bidding strategy, only the information before the time slot t and the predicted value of the time slot t can be observed. The load rebound of the subsequent time slot cannot be observed. From the simulation results, it can be seen that the algorithm can infer the load rebound through limited observation information, while tracking the load change well. The power deviation in all time slots can be maintained in a small range, thereby minimizing the loss of profit caused by power imbalance.

The observed price and consumer load of LSE in decision-making are predicted values, and the prediction error affects the decision-making effect. Therefore, we verify the performance of the algorithm in high uncertainty scenarios, as shown in Fig. 9.

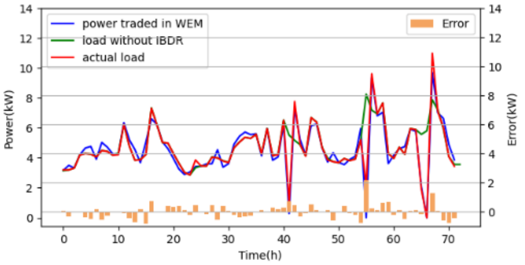


Fig. 9. Traded electricity in WEM and power consumption of consumers with high uncertainty

In the high uncertainty scenario, the maximum prediction error exceeds 20%. As can be seen from Fig. 9, the bidding strategy can still track the change of power curve well with high uncertainty. Although the power imbalance increases, it is maintained within an acceptable range.

Besides, since LSE often requires to interact with a large number of consumers in IBDR, the scalability of the algorithm also needs to be verified. We set up 500 consumers to verify the performance of the algorithm in the same scenario, as shown in Fig. 10.

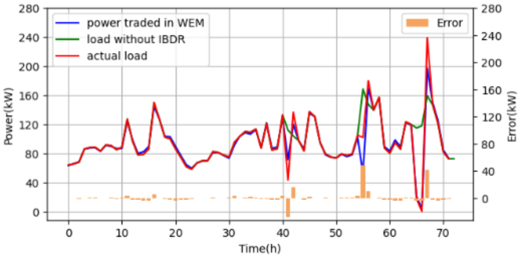


Fig. 10. Traded electricity in WEM and power consumption (500 consumers)

In the proposed algorithm, all the observations are the total value of consumers, e.g., total power, total response, etc. Therefore, the increase in the number of consumers does not cause additional burden on the algorithm. It can be seen from Fig. 10 that the algorithm can still maintain a relatively stable and good performance, indicating that the proposed algorithm can be adapted to the scene of a large number of consumers.

6 Conclusions and Further Works

In this paper, we propose an increment incentive mechanism for IBDR, in which consumers obtain incentive according to their incremental response. Through mathematical analysis, we find that the existing unified incentive mechanism is a special form of the increment incentive mechanism. By extending the unified incentive price to the IIF, consumer surplus can be distributed more reasonably. The excessive consumer surplus in the cumulative response is alleviated, and high-flexible consumer can obtain more profit by improving the response. The unit incentive cost of LSE is also reduced while increasing the response of consumers. Besides, the asynchronous double-interaction DRL algorithm has relatively stable and satisfactory performance in different scenarios in the simulation, which verifies that the proposed algorithm has satisfactory stability and adaptability, and it can deal with decision-making problems in a coupled environment well.

In future works, the performance of IIF with different forms for diversified consumer groups can be further studied, so as to achieve adaptive optimization of the IIF form according to the characteristics of consumer groups. In addition, in order to further improve system efficiency, the application of incremental incentive mechanisms can be expanded, e.g., price-based DR, peer-to-peer transactions, etc.

Acknowledgments

This work was sponsored by National Key Research and Development Program of China (Grant No. 2017YFE0132100).

References

[1] J. Shu, R. Guan, L. Wu and B. Han, “A bi-level approach for determining optimal dynamic retail electricity pricing of large industrial customers,” *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2267-2277, Mar. 2019.

[2] Z. Wang, H. Li, N. Deng, et al., “How to effectively implement an incentive-based residential electricity demand response policy? Experience from large-scale trials and matching questionnaires,” Energy Policy, vol. 141, pp. 111450, Apr. 2020.

[3] I. Antonopoulos, V. Robu, B. Couraud, D. Flynn, “Data-driven modelling of energy demand response behaviour based on a large-scale residential trial,” Energy and AI, vol. 4, pp. 100071, Apr. 2021.

[4] M. R. Sarker, M. A. Ortega-Vazquez and D. S. Kirschen, “Optimal Coordination and Scheduling of Demand Response via Monetary Incentives,” *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1341-1352, Dec. 2015.

[5] L. Wen, K. Zhou, J. Li, S. Wang, “Modified deep learning and reinforcement learning for an incentive-based demand response model,” Energy, vol. 205, pp. 118019, Jun. 2020.

[6] M. Rana, K. Rahi, T. Ray, R. Sarker, “An efficient optimization approach for flexibility provisioning in community microgrids with an incentive-based demand response scheme,” Sustainable Cities and Society, vol.74, pp. 103218, Aug. 2021.

[7] D. Muthirayan, D. Kalathil, K. Poolla and P. Varaiya, “Mechanism design for demand response programs,” IEEE Trans. Smart Grid, vol. 11, no. 1, pp. 61-73, Jan. 2020.

[8] Y. Chai, Y. Xiang, J. Liu, C. Gu, W. Zhang and W. Xu, “Incentive-based demand response model for maximizing benefit of electricity retailers,” J. Mod. Power Syst. Clean Energy, vol. 7, no. 6, pp. 1644-1650, Nov. 2019.

[9] Jindal, M. Singh and N. Kumar, “Consumption-aware data analytical demand response scheme for peak load reduction in smart grid,” IEEE Trans. Ind. Elec., vol. 65, no. 11, pp. 8993-9004, Nov. 2018.

[10] H. Aalami, H. Pashaei-Didani, S. Nojavan, “Deriving nonlinear models for incentive-based demand response programs,” *Int. J. Electr. Power Energy Syst.*, vol. 106, pp. 223-231, Oct. 2019.

[11] J. Lin, J. Dong, X. Dou, Y. Liu, P. Yang, T. Ma, “Psychological insights for incentive-based demand response incorporating battery energy storage systems: A two-loop Stackelberg game approach,” Energy, vol. 239, pp. 122192, Sep. 2022.

[12] M. Yu, S. H. Hong, Y. Ding and X. Ye, “An incentive-based demand response (dr) model considering composited DR resources,” IEEE Trans. Ind. Elec., vol. 66, no. 2, pp. 1488-1498, Feb. 2019.

[13] Jindal, M. Singh and N. Kumar, “Consumption-aware data analytical demand response scheme for peak load reduction in smart grid,” *IEEE Trans. Ind. Elec.*, vol. 65, no. 11, pp. 8993-9004, Nov. 2018.

[14] H. Yang, J. Zhang, J. Qiu, S. Zhang, M. Lai and Z. Y. Dong, “A practical pricing approach to smart grid demand response based on load classification,” *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 179-190, Jan. 2018.

[15] Y. Astriani, G. Shafiullah, F. Shahnia, “Incentive determination of a demand response program for microgrids,” Applied Energy, vol. 292, pp. 116624, Mar. 2021.

[16] Q. Hu, F. Li, X. Fang and L. Bai, “A framework of residential demand aggregation with financial incentives,” *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 497-505, Jan. 2018.

[17] M. Vanouni and N. Lu, “A reward allocation mechanism for thermostatically controlled loads participating in intra-hour ancillary services,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4209-4219, Sept. 2018.

[18] H. Zhong, L. Xie and Q. Xia, “Coupon incentive-based demand response: theory and case study,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1266-1276, May 2013.

[19] D. Qiu, D. Papadaskalopoulos, Y. Ye and G, Strbac, “Investigating the effects of demand flexibility on electricity retailers’ business through a tri-level optimisation model,” *IET Gener. Transm. Distrib*, vol. 14, no. 9, pp. 1739-1750, May, 2020.

[20] R. Henríquez, G. Wenzel, D. E. Olivares and M. Negrete-Pincetic, “Participation of demand response aggregators in electricity markets: optimal portfolio management,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4861-4871, Sept. 2018.

[21] E. Mahboubi-Moghaddam, M. Nayeripour, J. Aghaei, A. Khodaei and E. Waffenschmidt, “Interactive robust model for energy service providers integrating demand response programs in wholesale markets,” *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2681-2690, Jul. 2018.

[22] S. R. Konda, A. S. Al-Sumaiti, L. K. Panwar, B. K. Panigrahi and R. Kumar, “Impact of load profile on dynamic interactions between energy markets: a case study of power exchange and demand response exchange,” *IEEE Trans. Ind. Inform.*, vol. 15, no. 11, pp. 5855-5866, Nov. 2019.

[23] H. Xu, H. Sun, D. Nikovski, S. Kitamura, K. Mori and H. Hashimoto, “Deep reinforcement learning for joint bidding and pricing of load serving entity,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6366-6375, Nov. 2019.

[24] P. R. Thimmapuram and J. Kim, “Consumers’ Price Elasticity of Demand Modeling with Economic Effects on Electricity Markets Using an Agent-Based Model,” *IEEE Trans. on Smart Grid*, vol. 4, no. 1, pp. 390-397, March 2013.

[25] G. Li, Y. Huang and Z. Bie, “Reliability Evaluation of Smart Distribution Systems Considering Load Rebound Characteristics,” *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1713-1721, Oct. 2018.

[26] M. Muratori and G. Rizzoni, “Residential demand response: dynamic energy management and time-varying electricity pricing,” *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1108-1117, Mar. 2016.

[27] N. Mazzi, J. Kazempour and P. Pinson, “Price-taker offering strategy in electricity pay-as-bid markets,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2175-2183, Mar. 2018.

[28] Z. Baharlouei and M. Hashemi, “Efficiency-fairness trade-off in privacy-preserving autonomous demand side management,” *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 799-808, March 2014.

[29] Y. Li, C. Wang, G. Li, C. Chen, “Optimal scheduling of integrated demand response-enabled integrated energy systems with uncertain renewable generations: a stackelberg game approach,” *Energy Convers. Manag.*, Vol. 235, 2021.

[30] O. Jogunola et al., “Consensus algorithms and deep reinforcement learning in energy market: a review,” *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4211-4227, 15 March15, 2021.

[31] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. d. L. Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq, et al., "Deepmind control suite," *arXiv preprint arXiv*:1801.00690, 2018.

[32] S. Devlin, L. Yliniemi, D. Kudenko, and K. Tumer, “Potential-based difference rewards for multiagent reinforcement learning,” International Conference on Autonomous Agents and Multi-agent Systems, 2014.

[33] Q. Zang and L. Zhang, “Asymptotic behaviour of the trajectory fitting estimator for reflected Ornstein–Uhlenbeck processes,” *J. Theor. Probability*, vol. 3, pp. 1–19, 2017.

[34] Pecan Street Database. [Online]. Available: http://www.pecanstreet.org/.