Multi-Home Energy Coordination Scheduling Based on Adaptive Dynamic Programming

Kang Xiong and Qinglai Wei

***Abstract*—Implementing cooperative scheduling of multi-home microgrid energy and reducing the dependence on the main grid have become the focus of microgrid energy management research. This paper proposes a new multi-agent adaptive dynamic programming (MAADP) method for the cooperative control of distributed home energy. Each home is defined as a learning agent that needs to reasonably schedule the energy storage system to meet the respective load demand while accomplishing cooperative scheduling among the individual homes. In addition, an energy clearing center (ECC) is introduced to complete the energy exchange between each microgrid to protect the benefits of all parties. The proposed method adopts the learning strategy of "centralized learning and decentralized execution" to avoid the leakage of private information. The experimental comparison with the benchmark method verifies that the method can realize the cooperative scheduling of each home and reduce the dependence on the main grid.**

***Index Terms*—Multi-home scenarios, adaptive dynamic programming (ADP), optimal home energy management, smart grid**

I. Introduction

The global warming issue has strengthened people's environmental protection awareness, and using clean energy to replace fossil energy has become the central theme. With the maturity of photovoltaic (PV) and wind power generation technologies, home customers have widely used them. With the influx of clean energy into the main grid, the regular operation of the main grid is greatly affected. The emergence of the smart grid provides a solution to these problems. Smart grids can realize the rational dispatch of energy, improve energy utilization, and reduce energy waste. With the rise of artificial intelligence technology, more new technologies have been introduced into smart grids, including artificial neural networks (ANNs) [1, 2], fuzzy logic systems [3, 4], reinforcement learning (RL) [5, 6], and adaptive dynamic programming (ADP) [7−9]. This paper focuses on the problem of cooperative control of multiple home energy sources to reduce the dependence on the main grid.

The multi-microgrid energy scheduling problem [10] involves the joint scheduling of multiple homes or plants, which has a broader application than single-microgrid energy management [11] but is relatively more complex. Multi-microgrid energy management aims to jointly schedule energy from each microgrid to meet the load demand and reduce the dependence on the main grid. Existing methods of multi-microgrid energy management can be divided into three main categories: centralized, decentralized, and hybrid.

The centralized method (CM) [12] gives control of all microgrids to a central controller, which collects all information and then performs uniform training and scheduling. In Ref. [13], a centralized multi-objective optimization algorithm is proposed to achieve coordinated control among multiple microgrids and a balance between performance and cost finally. The CM can accomplish its goals well, but there are two main problems: the "curse of dimensionality" and privacy issues [14]. The decentralized method (DM) [15, 16] gives the decision-making power to each microgrid, and each microgrid performs energy scheduling with the local information it obtains. A decentralized distributed system is constructed using RL in Ref. [17], and a two-layer control strategy is used to realize the distributed control of each microgrid. In Ref. [18], a three-layer coordinate system control framework is proposed that can be used in grid-connected and islanded modes. The problem with DM is that it is difficult to achieve a balance. The emergence of hybrid method (HM) [19] alleviates the shortcomings of the above two methods, which combines centralized and decentralized methods in training and deployment to better balance the advantages and disadvantages. In Ref. [20], a multi-agent RL algorithm is proposed for secondary voltage control, which uses a "centralized learning and decentralized execution" learning strategy. In Ref. [21], a multi-agent RL algorithm is used to balance the benefits of microgrid participants, and a negotiating agent is introduced to solve the optimal equilibrium solution. In addition, the scalability of multi-home energy management system (MHEMS) is discussed in Ref. [22].

The above-mentioned multi-microgrid energy management scheme has achieved fruitful results, but some aspects can still be improved. Firstly, most existing schemes treat renewable energy and loads in one category and rarely consider individual homes. Secondly, most existing solutions still use discrete control strategies, which are less accurate than continuous control strategies. Thirdly, most of the solutions adopt offline learning methods, which makes it challenging to carry out practical applications.

Manuscript received: 26 April 2023; revised: 15 August 2023; accepted: 8 November 2023. *(Corresponding author: Qinglai Wei.)*

Citation: K. Xiong and Q. Wei, Multi-home energy coordination scheduling based on adaptive dynamic programming, *IJICS*, 2024, 29(1), 30–36.

Kang Xiong and Qinglai Wei are with Institute of Systems Engineering, Macau University of Science and Technology, Macao 999078, China (e-mail: 2109853wmi30001@student.must.edu.mo; qinglai.wei@ia.ac.cn).

Qinglai Wei is also with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and with School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

Digital Object Identifier 10.62678/IJICS202403.10101

This paper proposes a new multi-agent adaptive dynamic programming (MAADP) method for distributed control of home energy. The method introduces a "centralized learning and decentralized execution" learning strategy [23, 24] based on an action-dependent heuristic dynamic programming (ADHDP) method. The MAADP method is a data-based online learning method that learns in real-time interaction with the microgrid system. Each home is defined as a subsystem with learning capability. The acquired local information is used to complete the intelligent scheduling of the energy storage system (ESS) and achieve cooperative control of multiple homes. In addition, the MAADP method introduces an energy clearing center (ECC) for energy exchange between homes and centralized training.

II. Problem Formulation

In this section, we will introduce the multi-home energy management system, define the constraints of the relevant components, and present the optimized objectives.

A. Multi-Home Energy Management System

The structure of the MHEMS consists of three main components: the main grid, the home microgrid, and the ECC. Each home microgrid has its ESS, load, and renewable energy PV. ECC is responsible for energy clearing between individual homes, determining clearing prices, and completing centralized training. The ECC performs the energy-clearing process by purchasing energy from the main grid if there is not enough energy in the market to meet the load demand of all homes. Conversely, if energy exceeds demand, the excess energy is sold to the main grid.

The operation process of the MHEMS is as follows. First, at moment t, home microgrid i gets the electricity price of the main grid $C_{\mathrm{m}}(t)$, the electricity generated by PV $P_{\mathrm{pv}}^{i}(t)$, the load $P_{\mathrm{l}}^{i}(t)$, and the remaining capacity of the ESS $E_{\mathrm{b}}^{i}(t)$ to generate the charging and discharging power of the ESS $P_{\mathrm{b}}^{i}(t)$ and price of electricity for sale $C_{\mathrm{s}}^{i}(t)$. The energy exchanged $P_{\mathrm{e}}^{i}(t)$ can be calculated by Eq. (1), where $P_{\mathrm{e}}^{i}(t)>0$ means energy to be purchased, $P_{\mathrm{e}}^{i}(t)<0$ means energy to be sold, and $P_{\mathrm{e}}^{i}(t)=0$ means no energy transaction is required, and the sale price $C_{\mathrm{s}}^{i}(t)$ is valid only when energy is sold. Then, home microgrid i sends the energy exchange $P_{\mathrm{e}}^{i}(t)$ and the price $C_{\mathrm{s}}^{i}(t)$ to the ECC for energy trading. The ECC will set the final clearing price $C_{\mathrm{c}}(t)$ according to the set rules to complete the energy demand of each home.

$$P_{\mathrm{e}}^{i}(t)=P_{\mathrm{l}}^{i}(t)-P_{\mathrm{b}}^{i}(t)-P_{\mathrm{pv}}^{i}(t) \tag{1}$$

where the superscript i denotes the i-th home microgrid. $P_{\mathrm{b}}^{i}(t)>0$ means discharging, $P_{\mathrm{b}}^{i}(t)<0$ means charging, and $P_{\mathrm{b}}^{i}(t)=0$ is idle.

The ECC needs to follow the setup rules to set the final clearing price and the energy transactions between the individual home microgrids and the main grid. The rules are set as follows.

(1) Each home needs to provide an energy exchange P_{e}^{i} as well as a sale price C_{s}^{i}. The sale price is only valid when the energy is sold.

(2) The ECC needs to compare the effective selling price offered by each home microgrid with the electricity price of main grid and choose the smallest electricity price as the final clearing price.

(3) If the energy remaining in the ECC will be sold to the main grid after all load demands are met. The sale price is 0.8 times the price of electricity from the main grid. The benefits from the sale will be distributed to the individual customers according to the proportion of energy sold.

(4) If the ECC does not have enough energy, energy will be purchased from the main grid, and the cost will be shared in proportion to the home demand.

B. Component and Constraint

We use the standard battery model as ESS, and the battery model can be represented as Eq. (2).

$$E_{\mathrm{b}}^{i}(t+1)=E_{\mathrm{b}}^{i}(t)-P_{\mathrm{b}}^{i}(t)\times\eta_i(P_{\mathrm{b}}^{i}(t)) \tag{2}$$

where $\eta_i(\cdot)$ denotes the conversion efficiency of the battery.

The efficiency of the battery is derived in Ref. [25] as

$$\eta_i(P_{\mathrm{b}}^{i}(t))=0.898-0.173|P_{\mathrm{b}}^{i}(t)|/P_{\mathrm{rate}}^{i} \tag{3}$$

where $P_{\mathrm{rate}}^{i}>0$ is the rated power of the battery. The battery needs to satisfy certain constraints, which can be expressed as follows

$$\begin{gathered}P_{\mathrm{b,min}}^{i}\leqslant P_{\mathrm{b}}^{i}(t)\leqslant P_{\mathrm{b,max}}^{i},\\ E_{\mathrm{b,min}}^{i}\leqslant E_{\mathrm{b}}^{i}(t)\leqslant E_{\mathrm{b,max}}^{i}\end{gathered} \tag{4}$$

where $P_{\mathrm{b,min}}^{i}$ and $P_{\mathrm{b,max}}^{i}$ denote the battery's maximum charging power and maximum discharging power, respectively. $E_{\mathrm{b,min}}^{i}$ and $E_{\mathrm{b,max}}^{i}$ denote the battery's minimum capacity and maximum capacity, respectively.

The power generated by PV is related to the parameters of PV panels, area, real-time temperature, and solar radiation, and its maximum power generation is $P_{\mathrm{pv,max}}^{i}$. The maximum load demand for each home at each moment is $P_{\mathrm{l,max}}^{i}$. PV and load are provided with data. In addition, the selling price of home microgrid energy cannot exceed $C_{\mathrm{s,max}}^{i}$. The electricity price of the main grid cannot exceed $C_{\mathrm{m,max}}$. This paper uses data at 1 h intervals to validate and test the method.

C. Optimization Objective

The optimization objective is crucial to the training of the entire model. Each home microgrid has its decision-making system and optimization objective in the MHEMS. The optimization objective U_i of the home microgrid i is as follows

$$\begin{aligned}U_i=\sum_{t}^{\infty}\gamma_i^t\Bigg(&m_1^i\left(\frac{C_{\mathrm{cost}}^{i}(t)}{N_{\mathrm{cost}}^{i}}\right)^2+m_2^i\left(\frac{C_{\mathrm{batt}}^{i}(t)}{N_{\mathrm{batt}}^{i}}\right)^2+\\ &m_3^i\left(\frac{C_{\mathrm{sell}}^{i}(t)}{N_{\mathrm{sell}}^{i}}\right)^2\Bigg)\end{aligned} \tag{5}$$

where C_{cost}^{i} is the cost of purchasing or selling energy, C_{batt}^{i} is the penalty for the remaining capacity of the ESS, and C_{sell}^{i} is the penalty for selling price, respectively. Since the above three parts have different attributes, they must be divided by a weight to combine. N^i_{cost}, N^i_{batt}, and N^i_{sell} are the weights of the three parts, respectively. m^i_1, m^i_2, and m^i_3 are the proportions of the three parts, respectively. γ_i is the discount factor of the home microgrid. C^i_{cost}, C^i_{batt}, and C^i_{sell} are calculated as follows

$$C^i_{\text{cost}}(t) = \text{Cost}^i(t) \tag{6}$$

$$C^i_{\text{batt}}(t) = E^i_{\text{b}}(t) - P^i_{\text{b}}(t)\eta_i(P^i_{\text{b}}(t)) - (E^i_{\text{b,max}}/2) \tag{7}$$

$$C^i_{\text{sell}}(t) = C^i_{\text{s}}(t) \tag{8}$$

where $\text{Cost}^i(t)$ is the cost or revenue given by the ECC at t moment. The three weights are $P^i_{\text{l,max}} \cdot C_{\text{m,max}}$, $E^i_{\text{b,max}}/2$, and $C^i_{\text{s,max}}$, respectively. The overall objective U_c of the microgrid system can be expressed as follows

$$\min U_{\text{c}} = U_1 + U_2 + \cdots + U_H \tag{9}$$

where H is the number of homes. The MAADP method aims to achieve cooperative control among home microgrids and reduce the dependence on the main grid while safeguarding the benefits of all parties.

III. MAADP Method

In this section, we will first briefly introduce the ADHDP algorithm, then describe the whole structure of the MAADP method based on it, and finally, present the implementation details of the method.

A. Structure of MAADP

ADHDP is one of the most commonly used ADP algorithms. Werbos first proposed ADP, an intelligent optimization method with self-learning and adaptive functions, integrating optimal control, RL, and ANN. The ADHDP algorithm is an actor-critic method, with the actor being the policy network and the critic being the value network. The algorithm evaluates the policy a generated by the actor through the value q generated by the critic, and finally obtains the optimal policy by an iterative method. At moment t, the critic network is updated by minimizing the error δ_t, and the actor network is updated by minimizing value q_t by adjusting the policy a_t. The error δ_t is solved as follows

$$\delta_t = q_t - (r_t + \gamma q_{t+1}) \tag{10}$$

where r_t denotes the return at moment t and γ denotes the discount factor.

In the MHEMS, each home microgrid has an actor-critic network to have learning capabilities. In order to balance the benefits among home microgrids, the MAADP method introduces the learning strategy of "centralized learning and decentralized execution". Specifically, the MAADP method stores the actor network in each home microgrid, while all the critic networks are encapsulated in the ECC for centralized learning. ECC is only responsible for clearing energy and centralized training of the critic network of home microgrids.

The learning process of the MAADP method is as follows. First, at moment t, home microgrid i generates policy a^i_t and executes them based on the acquired state o^i_t. Then, the user sends the required energy exchange and selling price to the ECC for trading, which returns the user's cost and revenue, and the user calculates his return r^i_t based on the returned cost. In the next step, the home microgrid i sends the state o^i_t, the policy a^i_t, and the return r^i_t together to the ECC. Finally, all the critic networks in the ECC will use the uploaded information for learning. The respective critic networks of the home microgrid will feed a value q^i_t for learning the actor network. The overall network structure of MAADP is shown in Fig. 1.

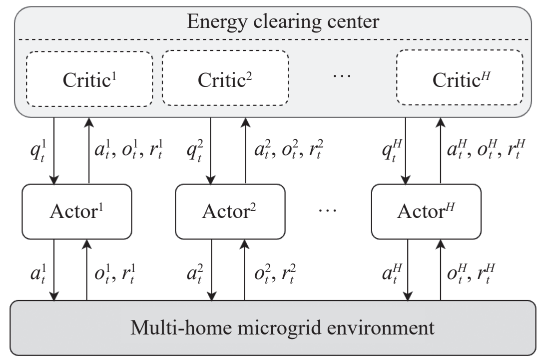

Figure 1 Network structure of MAADP.

The MAADP method is a data-based online learning method that learns from real-time interactions with multi-home microgrid systems. We add a rule system (RS) to the home microgrid to ensure that the actions generated by the actor network are feasible. We will introduce the rule system and the learning details in the rest of this section.

B. Rule System

Setting up the RS prevents each actor network from generating infeasible actions during the initial training phase. In the learning phase, after actor network of each home microgrid generates actions, the actor networks are processed by RS to replace the infeasible actions and guarantee the normal operation of the whole microgrid system. As learning proceeds, the probability of generating infeasible actions decreases, and the role of RS diminishes. The function of the RS in a home microgrid is primarily to ensure that the inputs and outputs of the ESS meet its capacity settings. All actor networks are built with outputs that meet the action limits, so there is no need for additional settings in the RS.

C. State and Action

The state o^i_t and the policy a^i_t of the home microgrid i at moment t can be expressed as

$$\begin{aligned} o^i_t &= \left\{P^i_{\text{pv}}(t), P^i_{\text{l}}(t), E^i_{\text{b}}(t), C_{\text{m}}(t)\right\}, \\ a^i_t &= \left\{P^i_{\text{b}}(t), C^i_{\text{s}}(t)\right\} \end{aligned} \tag{11}$$

All the critic networks are encapsulated in ECC, and they all have the same input state, including the states and policies of all home microgrids, and they output respective values q^i_t. The input state and output actions of the critic network at moment t, represented by $o^{\text{critic}^i}_t$ and $a^{\text{critic}^i}_t$, can be expressed as

$$o_t^{\text{critic}^i} = \left\{o_t^1, o_t^2, ..., o_t^H, a_t^1, a_t^2, ..., a_t^H\right\},\\ a_t^{\text{critic}^i} = \left\{q_t^i\right\} \tag{12}$$

The states and actions need to be pre-processed before they are fed into the neural network. All the state information needs to be mapped to the range $[-1, 1]$, and the actions need to be mapped to the range $[0, 1]$ to facilitate the processing by the neural network.

D. Network Structure and Network Update

Actor network of home microgrid uses a 4-10-2 network structure. Specifically, four-dimensional data input is processed by a hidden layer containing ten neurons, and two actions are output. A tanh activation function is used in the middle of the neural network structure. The first action is the charging and discharging power of the ESS, which is mapped to $[-1, 1]$ using the tanh activation function and then multiplied by $P^i_{\text{b,max}}$ to map the output to $[-P^i_{\text{b,max}}, P^i_{\text{b,max}}]$. The second action is the selling price, which is mapped to $[0, 1]$ using a sigmoid function and then multiplied by $C^i_{\text{s,max}}$ to map the output to $[0,\ C^i_{\text{s,max}}]$. All critic networks have the same structure, and the inputs are the states and actions of all home microgrids, which are processed by a hidden layer with 64 neurons to output values.

To alleviate the bootstrapping problem, we introduce two target networks actor_target and critic_target. The parameters of the two target networks are θ_i^- and w_i^-, respectively. The purpose of the target network has two main roles: Firstly, using the state of the next moment, actor_target predicts the action; secondly, based on the state and action of the next moment, critic_target predicts the value. The target network does not need to be trained and is updated using Formula (13). τ_i is the control weight of the old and new parameters, which is uniformly set to 0.8 in this paper.

$$\begin{cases} w_i^- \leftarrow \tau_i \cdot w_i + (1-\tau_i)\cdot w_i^-, \\ \theta_i^- \leftarrow \tau_i \cdot \theta_i + (1-\tau_i)\cdot \theta_i^- \end{cases} \tag{13}$$

To achieve better cooperative control of individual home microgrids, the overall return $U_{\text{c}}(t)$ solved by Eq. (9) is used as r_t^i to update the actor and critic networks.

IV. Computer Simulation

The MAADP method is validated, tested, and compared with the DM in this section. We use two home microgrids and one month of real data to validate our method. To simplify the experiment, we ignore the charging/discharging efficiency of the battery, but this does not affect the accuracy of the experiment. All experiments are run in a Python environment.

A. Comparison Method

We compare the MAADP method with the DM. In the DM, each home microgrid has its policy network and is trained individually. The DM uses the same network model as the MAADP method, except that the critic network is stored in each microgrid, and the other settings are the same as the MAADP method.

In our experiments, we analyze the scheduling strategies of the two methods, comparing the cost of home electricity consumption and the dependence on the main grid for both methods. In addition, we add a restriction on the selling price to the utility function to see the impact of the selling price on the cooperative schedule.

B. Experimental Data and Setup

The real data of a German city residence in November 2015 are extracted from the open source platform "Open Power System Data" (https://open-power-system-data.org). We use data from two residences as our experimental data. Due to the need for more data on real-time electricity prices, we use the electricity prices of one day of this month as the periodic electricity prices. The first 28 days of data are used for training, and the last two days are used for testing.

The two home microgrids are set up as Table 1. The maximum charge/discharge powers of ESS are the same, and the minimum loads are set to 0. During the experiments, the initial capacity of all ESSs is half of the maximum capacity. The discount factor is 0.998, and the learning rate of all networks is 0.002.

Table 1 Experimental setup of home.

i	$E^i_{\text{b,max}}$	$E^i_{\text{b,min}}$	$C_{\text{m,max}}$	$C^i_{\text{s,max}}$	$P^i_{\text{pv,max}}$	$P^i_{\text{l,max}}$	$P^i_{\text{b,max}}$
1	60	10	10	10	5	3	5
2	40	10	10	10	6	3	5

C. Experimental Comparison of Scenario One

In scenario one, we do not add a restriction on the selling price to the utility function to compare the strategies of the two methods.

Setup: The utility functions for the two home microgrids have the same parameters $m_1^i = 0.70$, $m_2^i = 0.30$, and $m_3^i = 0.00$. We use the first 28 days of data as training, with a total of 672 sets. The two comparison methods are trained for 50,000 iterations using 672 sets, and the last two days of data are used as a test.

Result: The test results of the MAADP method are shown in Fig. 2. Under the condition that there is no restriction on the selling price in the utility function, both homes set a selling price of around 5, as shown by the red polyline. At around the sixth moment, as shown by the blue bar, the load of home one requires more energy, while the load of home two requires less energy. In such a case, as shown by the dark blue polyline graph, home two sells a large amount of energy to the ECC, which is supplied to home one through the ECC, and the same happens at around moment 30. Home one similarly sells a small amount of energy to provide to home two at around moments 0 and 48. When the energy generated by the PV is higher, both homes choose to recharge their ESSs to maintain the remaining capacity at an intermediate level.

The test results of DM are shown in Fig. 3, and the selling prices of both homes are similar to those of the MAADP method. The black bar graphs indicate the charging and discharging of the ESS, from which it can be seen that the DM greatly reduces the use of the ESS in order to maintain the capacity of the ESS at an intermediate level. When PV

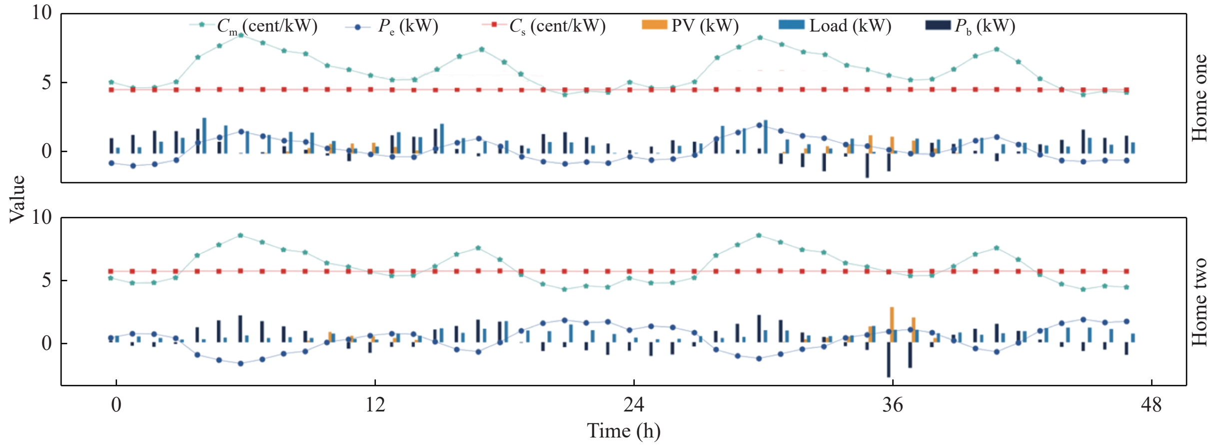

Figure 2 Test result of MAADP in scenario one.

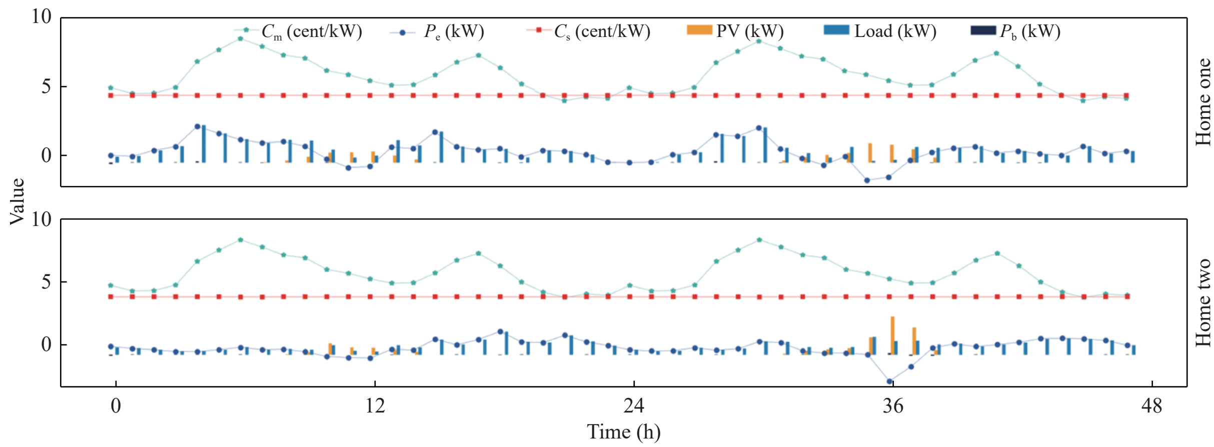

Figure 3 Test result of DM in scenario one.

generates more energy, both homes do not choose to charge the ESS but sell it to the ECC. At other times, both homes purchase energy from the ECC to meet the load demand.

In scenario one, both methods give similar selling prices but different strategies. Two homes in the MAADP method cooperate closely, and when one has a higher load demand, the other will sell its energy to reduce its dependence on the main grid. The test results of both methods are shown in Table 2. Evaluation metrics include two home costs, Cost^1 and Cost^2, average battery remaining capacity (ABRC), and energy purchased from the main grid. The DM is relatively poor in all comparison parts except for the average capacity of ESS, which is superior. The MAADP method has lower electricity costs for both homes and purchases less energy from the main grid.

D. Experimental Comparison of Scenario Two

In this section, we add a restriction on the selling price of the utility function. Specifically, we let the selling price be close to the price of the main grid to compare whether the strategies of the two methods change.

Setup: The parameters of the two home microgrid utility functions are $m_1^i = 0.70$, $m_2^i = 0.30$, and $m_3^i = 0.01$. Other settings are the same as those in scenario one.

Result: Due to the restrictions of the utility function, the selling prices of both homes are similar to the price of electricity from the main grid. Although the selling price is

Table 2 Comparison result in scenario one.

Algorithm	Cost^1 (cent)	Cost^2 (cent)	ABRC (kWh)	ABRC (kWh)	Purchased energy (kW)
DM	239.68	143.82	30.05	20.06	65.80
MAADP	99.11	53.12	15.96	11.70	28.17

Table 3 Comparison result in scenario two.

Algorithm	Cost^1 (cent)	Cost^2 (cent)	ABRC (kWh)	ABRC (kWh)	Purchased energy (kW)
DM	231.66	146.18	29.36	20.23	64.91
MAADP	127.16	28.49	16.13	11.62	28.81

changed, the strategies of the two home microgrids are the same as those in scenario one. The results of the comparison between the two methods are shown in Table 3. In the MAADP method, the cost of home one increases, while that of home two decreases. Because of the higher selling price, home two can earn more benefits. Home load in DM relies mainly on energy purchased from the main grid.

E. Experiment Summary

Our proposed MAADP method can accomplish the cooperative scheduling of multi-home energy in two experimental scenarios. Home one has more load demand, so home two will provide part of the energy to home one, reducing the dependence on main grid. When the sold price changes, the strategy remains unchanged. DM reduces the use of ESS in order to guarantee that the remaining capacity of ESS is maintained at the middle level, and its load demand mainly relies on the main grid. The experimental results show that MAADP can accomplish the cooperative scheduling of multi-home energy well and reduce the overall expense and reliance on the main grid.

V. Conclusion

This work describes and evaluates an MAADP energy management method for multi-home cooperative scheduling scenarios. The method introduces a learning strategy of "centralized learning and decentralized execution" based on the ADHDP method. The MAADP method reduces the cost of the entire microgrid and its dependence on the main grid by achieving cooperative scheduling of individual homes. In addition, we introduce an energy clearing center that can perform energy exchange between microgrids and centralized training of the value network of each home. Finally, the effectiveness and performance of our method are verified through comparative experiments. The shortcoming of the proposed method is that as the number of home microgrids increases, it increases the input state dimension of the value network, and the model training requires a more significant computational effort. In future work, we will consider more comprehensive microgrid models and address the problems of existing schemes.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (Nos. 2021YFE0206100 and 2018YFB1702300), the National Natural Science Foundation of China (No. 62073321), the National Defense Basic Scientific Research Program (No. JCKY2019203C029), and the Science and Technology Development Fund, Macao SAR (No. 0015/2020/AMJ).

References

[1] L. Haghnegahdar and Y. Wang, A whale optimization algorithm-trained artificial neural network for smart grid cyber intrusion detection, *Neural Comput. Appl*., 2020, 32(13), 9427–9441.

[2] M. Abdel-Nasser, K. Mahmoud, and H. Kashef, A novel smart grid state estimation method based on neural networks, *Int. J. Interact. Multimedia Artif. Intell*., 2018, 5(1), 92–100.

[3] S. M. Je and J. H. Huh, Estimation of future power consumption level in smart grid: Application of fuzzy logic and genetic algorithm on big data platform, *Int. J. Commun. Syst*., 2021, 34(2), e4056.

[4] M. G. Simões and A. Bubshait, Frequency support of smart grid using fuzzy logic-based controller for wind energy systems, *Energies*, 2019, 12(8), 1550.

[5] A. Kumari and S. Tanwar, A reinforcement-learning-based secure demand response scheme for smart grid system, *IEEE Internet Things J*., 2022, 9(3), 2180–2191.

[6] L. Zhang, Y. Gao, H. B. Zhu, and L. Tao, A distributed real-time pricing strategy based on reinforcement learning approach for smart grid, *Expert Syst. Appl*., 2022, 191, 116285.

[7] J. Yuan, S. S. Yu, G. D. Zhang, C. P. Lim, H. Trinh, and Y. Zhang, Design and HIL realization of an online adaptive dynamic programming approach for real-time economic operations of household energy systems, *IEEE Trans. Smart Grid*, 2022, 13(1), 330–341.

[8] Q. L. Wei, Z. H. Liao, R. Z. Song, P. J. Zhang, Z. Wang, and J. Xiao, Self-learning optimal control for ice-storage air conditioning systems via data-based adaptive dynamic programming, *IEEE Trans. Ind. Electron*., 2021, 68(4), 3599–3608.

[9] J. Yuan, G. D. Zhang, S. S. Yu, Z. Chen, Z. Li, and Y. Zhang, A multi-timescale smart grid energy management system based on adaptive dynamic programming and multi-NN fusion prediction method, *Knowl. -Based Syst*., 2022, 241, 108284.

[10] B. Zhou, J. T. Zou, C. Y. Chung, H. Z. Wang, N. Liu, N. Voropai, and D. S. Xu, Multi-microgrid energy management systems: Architecture, communication, and scheduling strategies, *J. Mod. Power Syst. Clean Energy*, 2021, 9(3), 463–476.

[11] D. R. Liu, Y. C. Xu, Q. L. Wei, and X. L. Liu, Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming, *IEEE/CAA J. Autom. Sin*., 2018, 5(1), 36–46.

[12] B. Sahoo, S. K. Routray, and P. K. Rout, A novel centralized energy management approach for power quality improvement, *Int. Trans. Electr. Energy Syst*., 2021, 31(10), e12582.

[13] A. Kargarian and M. Rahmani, Multi-microgrid energy systems operation incorporating distribution-interline power flow controller, *Electr. Power Syst. Res*., 2015, 129, 208–216.

[14] P. Kumar, Y. Lin, G. D. Bai, A. Paverd, J. S. Dong, and A. Martin, Smart grid metering networks: A survey on security, privacy and open research issues, *IEEE Commun. Surv. Tutor*., 2019, 21(3), 2886–2927.

[15] E. Samadi, A. Badri, and R. Ebrahimpour, Decentralized multi-agent based energy management of microgrid using reinforcement learning, *Int. J. Electr. Power Energy Syst*., 2020, 122, 106211.

[16] X. Q. Zhou, Q. Ai, and M. Yousif, Two kinds of decentralized robust economic dispatch framework combined distribution network and multi-microgrids, *Appl. Energy*, 2019, 253, 113588.

[17] C. Y. Guo, X. Wang, Y. H. Zheng, and F. Zhang, Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning, *Int. J. Electr. Power Energy Syst*., 2021, 131, 107048.

[18] L. W. Ju, Q. Zhang, Z. F. Tan, W. Wang, H. Xin, and Z. H. Zhang, Multi-agent-system-based coupling control optimization model for micro-grid group intelligent scheduling considering autonomy-cooperative operation strategy, *Energy*, 2018, 157, 1035–1052.

[19] W. Jiang, K. X. Yang, J. J. Yang, R. W. Mao, N. F. Xue, and Z. H. Zhuo, A multiagent-based hierarchical energy management strategy for maximization of renewable energy consumption in interconnected multi-microgrids, *IEEE Access*, 2019, 7, 169931–169945.

[20] D. Chen, K. A. Chen, Z. J. Li, T. S. Chu, R. Yao, F. Qiu, and K. X. Lin, Powernet: Multi-agent deep reinforcement learning for scalable powergrid control, *IEEE Trans. Power Syst*., 2022, 37(2), 1007–1017.

[21] X. H. Fang, J. K. Wang, G. R. Song, Y. H. Han, Q. Zhao, and Z. A. Cao, Multi-agent reinforcement learning approach for residential microgrid energy scheduling, *Energies*, 2019, 13(1), 123.

[22] F. Charbonnier, T. Morstyn, and M. D. McCulloch, Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility, *Appl. Energy*, 2022, 314, 118825.

[23] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, in *Proc. 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017, 6382–6393.

[24] J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, Counterfactual multi-agent policy gradients, in *Proc. 32nd AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, 2018, 2974–2982.

[25] T. Yau, L. N. Walker, H. L. Graham, A. Gupta, and R. Raithel, Effects of battery storage devices on power system dispatch, *IEEE Trans. Power Appar. Syst*., 1981, PAS-100, 375–383.

Kang Xiong received the MS degree in computer science and technology from Dalian University, Dalian, China, in 2021. He is currently pursuing the PhD degree in intelligent science and systems at Institute of Systems Engineering, Macau University of Science and Technology, Macao, China. His research interests include microgrid operation and energy management system.

Qinglai Wei received the BS degree in automation and the PhD degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2002 and 2009, respectively. From 2009 to 2011, he was a postdoctoral fellow at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently a professor at Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has authored three books and published over 70 international journal papers. His research interests include adaptive dynamic programming, neural networks based control, optimal control, nonlinear systems, and their industrial applications.