


Article

A Multi-Agent Deep-Reinforcement-Learning-Based Strategy for Safe Distributed Energy Resource Scheduling in Energy Hubs

Xi Zhang ^{1,2,*} , Qiong Wang ^{3,4}, Jie Yu ⁵, Qinghe Sun ¹, Heng Hu ¹ and Ximu Liu ⁵

¹ State Grid Smart Grid Research Institute, Beijing 102209, China; sunqinghe@geiri.sgcc.com.cn (Q.S.); huheng@geiri.sgcc.com.cn (H.H.)

² Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK

³ State Grid Beijing Municipal Electric Power Company, Beijing 100031, China; wangq23@mail.tsinghua.edu.cn

⁴ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

⁵ School of Electrical Engineering, Southeast University, Nanjing 210096, China; yujie@seu.edu.cn (J.Y.); 230238839@seu.edu.cn (X.L.)

* Correspondence: x.zhang14@imperial.ac.uk

Abstract: An energy hub (EH) provides an effective solution to the management of local integrated energy systems (IES), supporting the optimal dispatch and mutual conversion of distributed energy resources (DER) in multi-energy forms. However, the intrinsic stochasticity of renewable generation intensifies fluctuations in the system's energy production when integrated into large-scale grids and increases peak-to-valley differences in large-scale grid integration, leading to a significant reduction in the stability of the power grid. A distributed privacy-preserving energy scheduling method based on multi-agent deep reinforcement learning is presented for the EH cluster with renewable energy generation. Firstly, each EH is treated as an agent, transforming the energy scheduling problem into a Markov decision process. Secondly, the objective function is defined as minimizing the total economic cost while considering carbon trading costs, guiding the agents to make low-carbon decisions. Lastly, differential privacy protection is applied to sensitive data within the EH, where noise is introduced using energy storage systems to maintain the same gas and electricity purchases while blurring the original data. The experimental simulation results demonstrate that the agents are able to train and learn from environmental information, generating real-time optimized strategies to effectively handle the uncertainty of renewable energy. Furthermore, after the noise injection, the validity of the original data is compromised while ensuring the protection of sensitive information.

Keywords: privacy preservation; reinforcement learning; integrated energy system; distributed optimization; optimization algorithm



Citation: Zhang, X.; Wang, Q.; Yu, J.; Sun, Q.; Hu, H.; Liu, X. A Multi-Agent Deep-Reinforcement-Learning-Based Strategy for Safe Distributed Energy Resource Scheduling in Energy Hubs. *Electronics* **2023**, *12*, 4763. <https://doi.org/10.3390/electronics12234763>

Academic Editor: Fabio Mottola

Received: 31 October 2023

Revised: 18 November 2023

Accepted: 22 November 2023

Published: 24 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Energy hubs (EH) are characterized by the capability of integrating distributed renewable energy sources (RES), thereby facilitating a reduction in fossil fuel consumption and mitigation of carbon emissions [1–3]. However, due to the intrinsic stochasticity and variability in renewable energy, large-scale integration of wind and solar generation will widen the system's peak-to-valley difference. Additionally, extensive displacement of traditional fossil-fuel-based generators will result in a lack of system flexibility, thus driving substantial curtailment of RES [4–6]. Apart from the impacts of intermittent and uncertain RES, the stochastic nature of user loads, the diversity among energy sources, and the inter-dependencies among different energy forms also pose significant challenges regarding optimizing and managing energy systems [7,8]. In this context, model-based optimization approaches, e.g., mixed-integer linear programming (MILP) [9], dynamic programming [10,11], model predictive control (MPC), etc., have been widely used to address

such complex energy system scheduling problems. For instance, the authors in [12] employed MILP to optimize equipment selection and capacity configuration to minimize the annual economic cost. In [13], a two-stage MILP approach was employed to model the EH system, while Benders decomposition was utilized to solve the MILP problem. The authors in [14] applied dynamic programming methods to assess the performance of integrated electricity and heat networks by implementing decomposed electrical–hydraulic–thermal calculations for power flow. In [15], a weighted model predictive control energy scheduling regime was presented to enhance the resilience of EH clusters against contingencies. These optimization algorithms often rely on precise mathematical models and full information of parameters. However, due to the high complexity of EHs, establishing accurate models is challenging; meanwhile, the computational complexity of these algorithms grows exponentially with the increase in decision variables, limiting their applicability to large-scale EH optimization and scheduling problems. In this context, machine-learning (ML)-based algorithms are gaining popularity in recent years due to their low dependency on model accuracy and decent computational performance, and they have been applied in various fields, such as communication [16,17], bio-science [18,19], and energy [20,21]. As a type of commonly used ML approach, reinforcement learning (RL) has been widely used in energy system dispatch problems [22,23]. Through the interaction between agents and the environment, it learns the optimal action policy by trial and error to maximize cumulative rewards [24,25]. In the optimization and scheduling of EHs, RL can be employed to learn the dynamic characteristics of the system and the complex relationships between energy demand and supply, enabling autonomous decision making and optimization scheduling. In [26], the proximal policy optimization algorithm was used to address the dynamic scheduling problem of EHs in uncertain environments. However, this algorithm faces difficulties in convergence when dealing with non-stationary problems. In [27], the deep deterministic policy gradient (DDPG) method was applied to solve the EH energy management problem based on the Stackelberg game model, which can handle high-dimensional state and action spaces. However, it is unable to address large-scale EH cluster cooperation issues. In [28], a multi-agent deep-deterministic-policy-gradient (MADDPG)-based RL algorithm for optimizing EH clusters was employed to address the uncertainty of renewable energy generation. Compared to single-agent algorithms, the training process of this model exhibits improved stability and converging performance for tackling the interactions among multiple agents. Through the above analysis, existing energy scheduling methods, such as MILP, DDPG, and MADDPG, exhibit different characteristics in various aspects. MILP (mixed-integer linear programming)-based centralized optimization has been widely utilized in energy system scheduling problems; it is based on physical models and can provide the optimal solution. However, it requires full information from users, which is unrealistic when numerous end-user-side DERs are involved, as in our manuscript. Additionally, it highly depends on the accuracy of the mathematical model and is more computationally expensive to solve problems with a large number of users, thus being unsuitable for real-time scheduling [29,30]. RL-based algorithms such as DDPG-RL (deep deterministic policy gradient RL) are suitable for continuous state and action spaces, showing good adaptability to high-dimensional problems without the need for an accurate physical model, thus enabling offline training for energy system scheduling models and online application. Although showing good performance in handling continuous problems, DDPG is not suitable for multi-agent collaborative scenarios, which is the case in this paper [31,32]. Multi-agent deep reinforcement learning (MADRL) involves deep reinforcement learning in a multi-agent environment. As one of the commonly used MADRL methods, MADDPG employs centralized learning and decentralized execution, designed to address learning and decision making in multi-agent cases [33,34]. MADDPG is well-suited for multi-agent collaboration and adversarial scenarios, exhibiting strong adaptability with lax requirements for model precision in collaborative environments [16,35]. Despite the effectiveness of MADDPG algorithms in dealing with energy hub (EH) cluster scheduling problems in complex environments, they require access to sensitive information from subsystems,

which may pose potential risks of privacy leakage [36,37]. In this context, investigating how to ensure the accuracy and real-time performance of distributed subsystem scheduling while protecting data privacy warrants further exploration.

With the increasing number and variety of devices connected to the EH, a considerable amount of electricity consumption data is generated during the process of optimization scheduling [38]. These data may contain sensitive or private information from devices and users, posing significant security risks. Therefore, addressing privacy and security concerns in the optimization scheduling process is of utmost importance. In the data analysis and processing stage, commonly used privacy protection methods include homomorphic encryption (HE) [39,40] and differential privacy (DP) techniques. HE enables data analysis and processing while preserving the confidentiality of plaintext data [41]. In [42], HE algorithms have been applied to address privacy protection challenges in distributed energy management frameworks. However, HE algorithms suffer from high computational complexity, leading to higher resource utilization, degraded system performance, and higher costs. On the other hand, DP methods involve simple operations, such as data perturbation and noise injection [43]. These techniques are more suitable for privacy protection tasks in large-scale EH clusters, particularly under limited computational capabilities and resource constraints. However, excessive introduction of noise, while it greatly enhances the privacy protection of sensitive data, can lead to a decline in the performance of the EH network and instability in control [44]. Therefore, the trade-off between privacy protection and the performance of energy system dispatch is significant.

Based on the aforementioned discussions, algorithms combining DP and MADDPG (denoted by DP-MADDPG) are worth investigating for solving EH cluster scheduling problems with data privacy protection. Different from approaches that combine MADDPG and HE, which require more computational resources, thus squeezing communication and computation resource allocation for other tasks within the system, DP-MADDPG exhibits low computational complexity, requiring lower computational resources but achieving decent privacy protection performance, especially in scenarios involving multiple agents. The comparisons of different algorithms are demonstrated in Table 1. In [45], DP-MADDPG has been applied to address the issue of optimal power scheduling for microgrids with data privacy protection. However, it does not consider the complex interactions between different energy carriers in multi-energy systems. In this paper, we extend the utilization of this approach to a heat–electricity–gas system in EHs, aiming to effectively solve the optimization problem in EH clusters.

Table 1. Comparative analysis of different energy scheduling methods.

Algorithm	MILP	DDPG	MADDPG	HE-MADDPG	DP-MADDPG
Feature					
Lax requirement for model accuracy		✓	✓	✓	✓
Adaptability to multi-agents			✓	✓	✓
Advantage in convergence speed		✓	✓		✓
Privacy protection				✓	✓

✓ indicates the algorithm is characterised by the corresponding feature.

The main contributions of this paper are summarized as follows:

- (1) The DP-MADDPG algorithm is adopted for distributed management of the EH cluster system. Each agent independently controls the operation of its local system and adjusts its local policy based on real-time observations and reward signals, enhancing the robustness and reliability of the scheduling decisions. Furthermore, through

collaboration among multiple agents, the method addresses complex scheduling issues and improves the energy utilization efficiency of the system.

- (2) Data privacy concerns are effectively addressed using the presented method. This method dynamically introduces noise interference and utilizes an energy storage system (ESS) to attenuate noise, ensuring the preservation of external transaction data while perturbing internal network data. Additionally, an effective evaluation mechanism for EH privacy protection is established to mitigate the impact of data correlation on evaluation results, enabling intelligent agents to generate noise data that satisfy the constraint conditions within a reasonable range.

The rest of this paper is organized as follows. Section 2 presents the model structure and equipment type for EH clusters. EH’s optimal scheduling approach is detailed in Section 3. In Section 4, simulation results are presented to show the performance of the proposed approach, whilst the conclusion is drawn in Section 5.

2. Integrated Energy System Structure and Equipment Model

2.1. Integrated Energy System Structure

Figure 1 illustrates that EH comprises power grids, district heating networks, and gas networks. Through the utilization of diverse energy conversion and storage devices, EH facilitates the mutual supplementation and efficient utilization of energy resources among these networks, meeting diverse load requirements. Additionally, EH mitigates the challenge of inadequate electricity generation resulting from the unpredictable fluctuations of renewable energy sources by procuring energy from the distribution grid and natural gas networks.

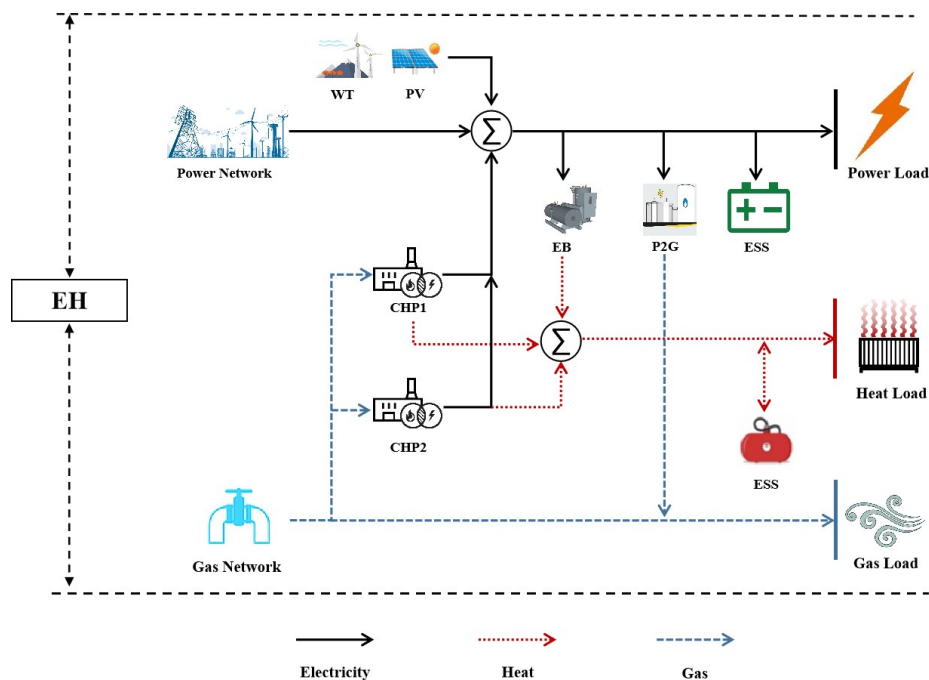


Figure 1. The framework of EH.

2.2. Model of Devices

- (1) Combined heat and power (CHP): CHP is an efficient energy utilization system that achieves integrated utilization of energy by simultaneously generating electricity and heat through the combustion of natural gas. The model can be described as follows:

$$P_{n,i,t}^{CHP} = \eta_e^{CHP} \times G_{n,i,t}^{CHP} \tag{1}$$

$$H_{n,i,t}^{CHP} = \eta_h^{CHP} \times G_{n,i,t}^{CHP} \tag{2}$$

- (2) Electric boiler (EB): EB is a device that converts electrical energy into thermal energy, used to meet the heat network’s load requirements when CHP is not operational.

$$H_{i,t}^{EB} = \eta^{EB} \times P_{i,t}^{EB} \tag{3}$$

- (3) Power to Gas (P2G): P2G is an energy conversion technology. By converting surplus electricity into natural gas, P2G systems can store energy in the natural gas grid to meet peak energy demands or periods when renewable energy generation falls short. This helps alleviate the challenges posed by the intermittency and fluctuations in renewable energy sources. The model is as follows:

$$G_{i,t}^{P2G} = \eta^{P2G} \times P_{i,t}^{P2G} \tag{4}$$

- (4) Energy storage model: Energy storage unit system is utilized for balancing load and supply in a network. The output model can be described as follows:

$$E_{i,t+1}^X = (1 - \alpha_x^{ES})E_{i,t}^X + (P_{i,t}^{X,ch}\eta_x^{ch} - \frac{P_{i,t}^{X,dis}}{\eta_x^{dis}}\Delta t), \tag{5}$$

where x denotes the type of energy storage and e , h , and g denote the grid, heat, and gas networks, respectively.

It should be emphasized that network losses also have significant impacts on the decisions of economic energy dispatch. Since the networks for different energy carriers are not explicitly modeled in this paper, the impacts of network losses are considered in the energy prices.

2.3. Constraints

- (1) Energy Balance Constraint: The power balance constraint of the entire integrated energy system is expressed as

$$P_{i,t}^{net} + P_{i,t}^{PV} + P_{i,t}^{WT} + \sum_{n=1}^N P_{n,i,t}^{CHP} - P_{i,t}^{EB} - P_{i,t}^{P2G} + P_{i,t}^{e,dis} - P_{i,t}^{e,ch} = L_{i,t}^e \tag{6}$$

$$\sum_{n=1}^N H_{n,i,t}^{CHP} + H_{i,t}^{EB} + P_{i,t}^{h,dis} - P_{i,t}^{h,ch} = L_{i,t}^h \tag{7}$$

$$G_{i,t}^{net} + G_{i,t}^{P2G} - \sum_{n=1}^N G_{n,i,t}^{CHP} = L_{i,t}^g \tag{8}$$

- (2) Equipment Operating Constraints: For P2G, CHP, and EB devices, power constraints and ramp constraints must be adhered to during operation as follows:

$$P_{min}^{P2G} \leq P_{i,j}^{P2G} \leq P_{max}^{P2G} \tag{9}$$

$$|\Delta P_{i,t}^{P2G}| \leq P_{ramp}^{P2G} \tag{10}$$

$$P_{min}^{EB} \leq P_{i,j}^{EB} \leq P_{max}^{EB} \tag{11}$$

$$|\Delta P_{i,t}^{EB}| \leq P_{ramp}^{P2G} \tag{12}$$

$$G_{min}^{CHP} \leq G_{i,j}^{CHP} \leq G_{max}^{CHP} \tag{13}$$

$$|\Delta G_{n,i,t}^{CHP}| \leq G_{ramp}^{CHP} \tag{14}$$

- (3) Energy Storage Device Constraints: The capacity constraints and ramp constraints to be satisfied by energy storage devices on different networks are expressed as the following equations:

$$E_{min}^X \leq E_{i,j}^X \leq E_{max}^X \tag{15}$$

$$I_{i,t}^{X,ch} + I_{i,t}^{X,dis} \leq 1 \tag{16}$$

$$P_{min}^{X,ch} I_{i,t}^{X,ch} \leq P_{i,t}^{X,ch} \leq P_{max}^{X,ch} I_{i,t}^{X,ch} \tag{17}$$

$$P_{min}^{X,dis} I_{i,t}^{X,dis} \leq P_{i,t}^{X,dis} \leq P_{max}^{X,dis} I_{i,t}^{X,dis}, \tag{18}$$

where the non-negative integer $I_{i,t}^{X,ch}$ and $I_{i,t}^{X,dis}$ are introduced to ensure that charging and discharging behaviors do not occur simultaneously.

In this paper, we are focused on the scenario of energy hubs, covering a couple energy conversion infrastructures, e.g., combined heat and power (CHP), electric boilers, electrolyzers, various forms of storage, which enable the interchange of energy forms between heat, electricity, and hydrogen and natural gas. However, some of the energy infrastructures are geographically specific and highly rely on regulatory environments; for example, CHP is commonly used in regions at high latitude but barely seen in southern areas. Under the carbon neutrality target, policies are coming up to facilitate the displacement of gas boilers by electric heaters (heat pumps or electric boilers), but there is still a long way to go before large-scale heating system electrification. Regarding P2G devices, affordability and security issues still highly restrict their large-scale deployment, which highly depends on technology advancement and political incentives. However, the decarbonisation pathway proposed by different countries may vary tremendously, which drives the preference of one type of energy infrastructure and suppresses the development of another. To this end, it is important to stress that the scenarios in this paper are not omni-applicable under different conditions; however, the proposed method provides meaningful insight regarding solving economic energy dispatch across multi-agents; the types of resources and application scenarios can be changed accordingly.

2.4. Carbon Trading Cost Model

Ignoring the carbon emissions from renewable power generators and energy storage, devices participating in carbon trading are CHP and P2G devices. For each carbon emission source, if the actual carbon emissions exceed the allocated carbon quota obtained for free, the excess portion needs to be purchased in the carbon trading market. The remaining quota can be sold. Therefore, the carbon trading cost model can be established as follows:

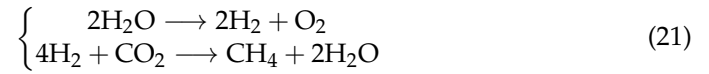
$$C_{i,t}^{CO_2} = C_{i,t}^{CHP} + C_{i,t}^{P2G} \tag{19}$$

- (1) CHP Carbon Trading Cost: CHP units are one of the main carbon emission sources in the energy system. Assuming that the total carbon emission intensity and quota are proportional to the actual output, the carbon-related cost can be calculated as follows:

$$C_{i,t}^{CHP} = \pi_t^{CO_2} (E^{CHP} - e^{CHP}) \sum_{n=1}^N (P_{n,i,t}^{CHP} + H_{n,i,t}^{CHP}) \tag{20}$$

- (2) P2G Carbon Trading Cost: The P2G unit can capture CO₂ from power plants or biogas. As shown in Equation (21), the conversion process of P2G can be divided into two

steps: electrolytic hydrogen production and methanation, where the volume of CO₂ consumed in this process is equal to the volume of CH₄ produced.



Therefore, the output of the P2G unit can be converted into an equivalent volume of CH₄, allowing us to further determine the reduction in carbon emission intensity achieved by the P2G unit.

$$E^{\text{P2G}} = \rho^{\text{CO}_2} V^{\text{CO}_2} = \rho^{\text{CO}_2} \frac{G_{i,t}^{\text{P2G}}}{\alpha^{\text{CH}_4}} \quad (22)$$

In this context, $\rho^{\text{CO}_2} = 1.977 \text{ kg/m}^3$ represents the gas density of CO₂, and α^{CH_4} denotes the calorific value of natural gas CH₄, which takes the value of 9.87 kWh/m³. Since the P2G unit is not a carbon emission source, its carbon quota is set to zero. Thus, the calculation of carbon trading costs can be represented as follows:

$$C_{i,t}^{\text{P2G}} = -\pi_t^{\text{CO}_2} E^{\text{P2G}} G_{i,t}^{\text{P2G}} \quad (23)$$

2.5. Objective Function

Minimizing the total operating cost of the integrated energy system is chosen as the objective function, which includes costs associated with external energy procurement, equipment operation, and maintenance, as well as carbon trading. The specific calculation method is as follows:

$$C_{i,t}^{\text{buy}} = \pi_t^e P_{i,t}^{\text{net}} + \pi_t^g C_{i,t}^{\text{net}} \quad (24)$$

$$C_{i,t}^{\text{oper}} = \pi^{\text{CHP}} \sum_{n=1}^N (P_{n,i,t}^{\text{CHP}} + H_{n,i,t}^{\text{CHP}}) + \pi^{\text{EB}} H_{i,t}^{\text{EB}} + \pi_t^{\text{ES}} (P_{i,t}^{\text{x,ch}} + P_{i,t}^{\text{x,dis}}) + \pi^{\text{P2G}} G_{i,t}^{\text{P2G}} \quad (25)$$

Based on the above discussion, the optimization scheduling problem described in this paper can be formulated as follows:

$$\begin{aligned} \min \sum_{i=1}^M (C_{i,t}^{\text{buy}} + C_{i,t}^{\text{oper}} + C_{i,t}^{\text{CO}_2}) \\ \text{s.t. (1) - (25)} \end{aligned} \quad (26)$$

3. A Real-Time Optimal Energy Scheduling Method for EH Based on Distributed Deep Reinforcement Learning

In this section, each complete and independent EH is regarded as an agent responsible for controlling energy dispatch operations within the system. The optimization of dispatch problems is formulated as a Markov decision process (MDP), and the global optimal decision is obtained through experience sharing and collaborative training. Due to the fluctuation and uncertainty in renewable energy output and load demand in the environment of EH, as well as the involvement of multiple variables, such as different energy sources, loads, devices, and markets, the combination of state space and action space exhibits explosive growth. To address these challenges, the MADDPG algorithm is adopted, which excels in handling complex tasks with high-dimensional state and action spaces while employing an adaptive strategy to cope with environmental uncertainties.

3.1. MADDPG Algorithm

Traditional algorithms like deep Q-learning and DDPG often encounter issues such as unstable training and convergence difficulties when dealing with non-stationarity in multi-agent environments. To address these challenges, the MADDPG algorithm has

been developed as a deep RL approach based on the deterministic policy gradient (DPG) and actor–critic framework, specifically designed for multi-agent settings. In MADDPG, each individual agent maintains its own actor and critic network, responsible for learning policies and evaluating policy value functions, respectively. During the training process, agents interact with the environment, selecting actions based on their actor networks and receiving rewards and subsequent states [46–48]. The experiences of the agents are stored in a shared experience replay buffer. When updating network parameters, agents sample data from the experience replay buffer, calculate gradients, and update their respective network parameters accordingly. Furthermore, MADDPG incorporates techniques such as target networks and random sampling from the experience replay buffer to enhance training stability. These mechanisms contribute to the effectiveness of MADDPG in addressing the optimization problem of energy scheduling. Detailed descriptions of the algorithm’s design and practical applications will be provided in the subsequent section.

3.2. Parameter Space

In traditional reinforcement learning, the MDP describes the interactive process between a single agent and the environment, where the agent selects actions based on the current state and evaluates the quality of its behavior through reward signals. MADDPG can be regarded as an extension of MDP for multi-agent scenarios. Thus, a reinforcement learning model for integrated energy systems can be represented by three essential components: the state space S_i , action space A_i , and reward space R_i of agent i .

- (1) State space: At time slot t , the state space of an EH cluster primarily encompasses the renewable energy generation (including wind power and photovoltaic generation) within each agent’s region, the load of the three energy networks, the gas consumption of CHP units, the electricity consumption of EB and P2G devices, the electricity price, gas price, and the charging and discharging actions of energy storage systems. It can be defined as follows:

$$s_{i,t} = \{t, G_{n,i,t}^{CHP}, P_{i,t}^{CHP}, P_{i,t}^{WT}, L_{i,t}^e, L_{i,t}^h, L_{i,t}^s, P_{i,t}^{EB}, P_{i,t}^{P2G}, \pi_t^e, \pi_t^g\} \tag{27}$$

with $s_{i,t} \in S_i$.

- (2) Action space: The action space variables mainly include controllable energy conversion devices and energy storage devices, which can be indicated as follows:

$$a_{i,t} = \{\Delta P_{n,i,t}^{CHP}, \Delta P_{i,t}^{P2G}, \Delta P_{i,t}^{EB}, P_{i,t}^{e,dis} / P_{i,t}^{e,ch}\} \tag{28}$$

with $a_{i,t} \in A_i$.

- (3) Reward function: The reward of agent i on given state s_i , t , and action a_i can be described as

$$r_{i,t} = \begin{cases} \gamma(C_{i,t}^{buy} + C_{i,t}^{oper} + C_{i,t}^{CO_2}), & \lambda = 0 \\ -\zeta\lambda, & \text{otherwise.} \end{cases} \tag{29}$$

where the λ is an integer parameter, indicating the number that does not satisfy (6)–(18) at time slot t .

- (4) Algorithm chart: The optimal energy scheduling process for EH cluster based on MADDPG is shown in Algorithm 1.

The essence of this approach is to search the optimal solution in the feasibility region defined by the optimization problem. Instead of using the traditional MILP method, which can be computationally prohibitive for real-time application, we turn to the MADDPG method for offline training. To achieve this, we first established the integrated heat–electricity–gas system model, where the operational characteristics of various energy conversion components and their interactions are specifically taken into account. Then, we formulated the EH economic dispatch optimization problem, which serves as the environment of the MADDPG model. The reward of the model includes two parts: (1) the revenue

gain (the objective of the optimization problem), and (2) the constraint violation penalty. Note that the coefficient of the constraint violation penalty is set very large so that, once an action violates a constraint, the reward will be particularly bad. In this way, this model learns offline to search the solutions with the highest reward, which can be interpreted as taking the actions that can lead to the highest revenue without violating physical constraints of the energy system. Given real-time EH operational conditions (environment parameters), the trained model tends to provide a decent decision for economic energy dispatch.

Algorithm 1: Distributed Energy Management by MADDPG.

Input: $G_{n,i,t}^{CHP}, P_{i,t}^{PV}, P_{i,t}^{WT}, L_{i,t}^e, L_{i,t}^h, L_{i,t}^g, P_{i,t}^{EB}, P_{i,t}^{P2G}, \pi_t^e, \pi_t^g$
Output: $\Delta P_{n,i,t}^{CHP}, \Delta P_{i,t}^{P2G}, \Delta P_{i,t}^{EB}, P_{i,t}^{e,dis} / P_{i,t}^{e,ch}$

```

1 for episode = 1 to  $\mathcal{H}$  do
2   Initialize  $G_{n,i,t}^{CHP}, P_{i,t}^{EB}, P_{i,t}^{P2G}, E_{i,t}^h, E_{i,t}^e$ 
3   for stept = 1 to  $\mathcal{T}$  do
4     Each agent obtains environmental parameters.
5     Actor network output action  $\Delta G_{n,i,t}^{CHP}, \Delta P_{i,t}^{P2G}, \Delta P_{i,t}^{EB}, P_{i,t}^{e,dis} / P_{i,t}^{e,ch}$ 
6     According to (6)–(18), and calculate the  $r_{i,t}$  value by Equation (29).
7     Update the state  $s_{i,t+\Delta t}$ .
8     if episode  $\geq \mathcal{H}'$  then
9       | Agent learns by extracting historical data from a pool of experience
10    end
11    Store  $[s_{i,t}, a_{i,t}, s_{i,t+\Delta t}, r_{i,t}]$  into the experience pool
12  end
13 end

```

3.3. EH Privacy Protection Based on Differential Privacy

In the utilization of reinforcement learning for training optimization scheduling models, the interaction between agents and the environment gives rise to security risks associated with data privacy breaches. Particularly, when agents engage in data transactions with external power and gas grids, the internal parameters of the agents become more susceptible to leakage. In order to safeguard data privacy in EH, we adopt an efficient and computationally simple approach known as local differential privacy [49–54]. This approach not only allows for quantifying the strength of privacy protection but also enables the application of the noise addition process at each EH node. By individually adding noise to the local privacy information of each agent, the probability of privacy leakage is greatly reduced.

Algorithm Chart

The optimal energy scheduling process for EH cluster based on MADDPG is shown in Algorithm 1.

We employ the Laplace mechanism to add noise to the data as a privacy-preserving measure, and each agent is responsible for controlling this noise addition process. Specifically, a local privacy dataset of agent i is first introduced and expressed as

$$\mathcal{D}_{i,t} = \{G_{n,i,t}^{CHP}, P_{i,t}^{PV}, P_{i,t}^{WT}, L_{i,t}^e, L_{i,t}^h, L_{i,t}^g, P_{i,t}^{EB}, P_{i,t}^{P2G}, P_{i,t}^{e,dis} / P_{i,t}^{e,ch}, P_{i,t}^{h,dis} / P_{i,t}^{h,ch}\} \quad (30)$$

Secondly, the dataset would be mapped into $x_{i,t} = f(\mathcal{D}_{i,t}) \in \mathbb{R}^d$ and used to generate the Laplace noise $Lap_n(\frac{\Delta f}{\epsilon})$ to construct DP vector denoted as

$$y_{i,t} = f(\mathcal{D}_{i,t}) + (Lap_1(\frac{\Delta f}{\epsilon}), \dots, Lap_d(\frac{\Delta f}{\epsilon}))^T, \quad (31)$$

where Δf and ϵ are the sensitivity and privacy budget of function f , respectively.

The privacy protection efficiency of agent i is assessed by computing the discrepancy between the original privacy information $x_{i,t}$ and the perturbation information $y_{i,t}$. The formula is defined as follows:

$$\sigma_{i,t} = \frac{\sqrt{(x_{i,t} - y_{i,t})^T S_{i,t}^{-1} (x_{i,t} - y_{i,t})}}{\|x_{i,t}\|_2} \tag{32}$$

where $S_{i,t}$ is denoted as the covariance matrix. Simultaneously, by incorporating the constraints (6)–(18), the agent is guided to select noise addition actions that not only satisfy the constraints but also achieve the desired level of privacy protection effectiveness.

Due to the negative impact of introducing external noise on the stability, security, and reliability of energy networks, utilizing internal ESS within the EH to provide the required additional energy for noise addition can mitigate this effect. By transforming the source of noise to be internal to the network, the impact of introducing noise can be effectively managed. This enables flexible adjustment and control of noise introduction according to the specific requirements and operational states of the network. Additionally, the ESS plays a role in energy balancing within the network, ensuring that sensitive data within the EH are perturbed without affecting energy transactions between the EH and external sources. Consequently, the added noise can be defined as

$$\sum \text{Lap}_n\left(\frac{\Delta f}{\epsilon}\right) = \sum_{k \in \mathcal{K}} E_{ik,t}^{X,\text{noise}}, \tag{33}$$

where the energy obtained through the ESS and used as noise $E_{ik,t}^{X,\text{noise}}$ follows the Laplace distribution with the following probability density function,

$$f(x|\mu, \lambda) = \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}}, \tag{34}$$

where μ usually takes 0, and $\lambda = \Delta f/\epsilon$.

4. Case Studies

The presented optimization scheduling model is applied to a cluster comprising four distributed EHs. Each EH can achieve a supply–demand balance through electricity and gas procurement operations. Time-of-use pricing is adopted for electricity and gas procurement/sales from the grid, with differentiated prices during different time intervals, as depicted in Figure 2. The operational parameters of the EH’s devices are provided in Table 1.

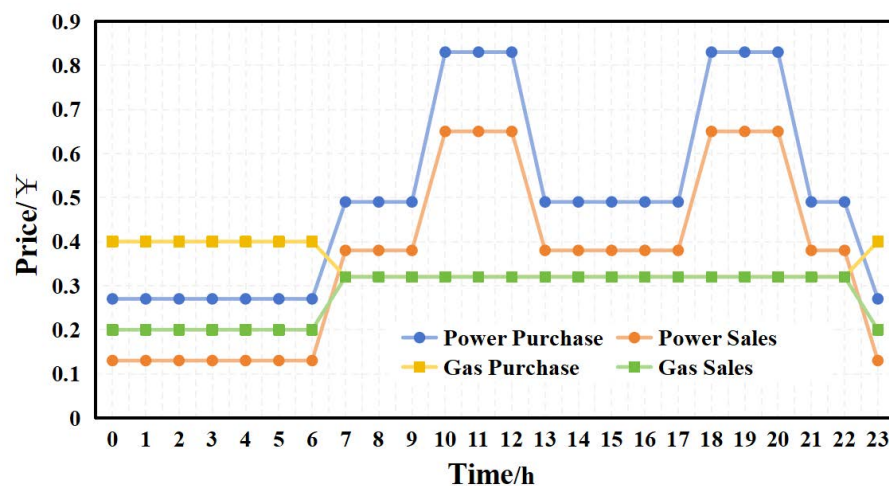


Figure 2. Electricity and gas prices.

4.1. Analysis of Optimized Schedule Results

To address the aforementioned cluster, this paper establishes an EH cluster optimization scheduling model based on the DP-MADDPG algorithm and conducts simulation experiments in the Python 3.7 environment to validate the proposed methodology’s effectiveness. The DP-MADDPG algorithm’s specific parameters are outlined in Tables 2 and 3. The model consists of four agents, and Agent-4 is chosen as a case study for the experimental analysis. In this paper, all the models are trained for 2500 episodes. By summing up the rewards obtained by each agent across 24 time steps in every round, the total reward for each iteration is calculated, and the data are averaged every 50 cycles. The convergence of the reward value for Agent-4 and the total reward value are illustrated in Figure 3.

Table 2. Energy conversion devices and energy storage unit parameters.

$P_{n,i}^{CHP}$	$P_{max}^{CHP}/P_{min}^{CHP}$ [kW]	P_{ramp}^{CHP} [kW]	Initial $P_{n,i}^{CHP}$ [kW]	$\eta_h^{CHP}/\eta_e^{CHP}$
$P_{1,i}^{CHP}$	150/0	60	45	0.5/0.4
$P_{2,i}^{CHP}$	110/0	45	36	0.5/0.4
	$P_{max}^{EB}/P_{min}^{EB}$ [kW]	P_{ramp}^{EB} [kW]	Initial $P_{n,i}^{EB}$ [kW]	η_e^{EB}
P_i^{EB}	100/0	15	45	0.9025
	$P_{max}^{P2G}/P_{min}^{P2G}$ [kW]	P_{ramp}^{P2G} [kW]	Initial $P_{n,i}^{P2G}$ [kW]	η_e^{P2G}
P_i^{P2G}	150/0	24	56	0.83
ESS	E_{max}^X/E_{min}^X [kWh]	Initial E_i^X [kWh]	α_x^{ES}	η_x^{ch}/η_x^{dis}
E_i^e	1000/−1000	0	0.01	0.95/0.95
E_i^h	800/−800	0	0.01	0.95/0.95

Table 3. Model-specific parameter settings.

Parameters	Critic	Actor
Learning rate	0.0001	0.001
Soft update coefficient	0.01	0.01
Number of layers of neural network	2	2
Number of neural per layer	64	64
Activation function of hidden layer	Relu	Relu
Activation function of output layer	/	Tanh
Number of episodes	10,000	10,000
The number of times per episode	24	24
Size of experience replay unit	100,000	100,000

Based on the graph, it can be observed that, during the initial stages of training, when the action networks are in the exploration phase, the reward values are initially low and exhibit significant fluctuations. However, as the intelligent agents begin to learn from historical data extracted from the experience replay buffer, the reward values gradually show a clear upward trend. Around 500 episodes, the bonus value curve stabilizes and stays at a higher level. Eventually, the total reward value converges to around −4000.

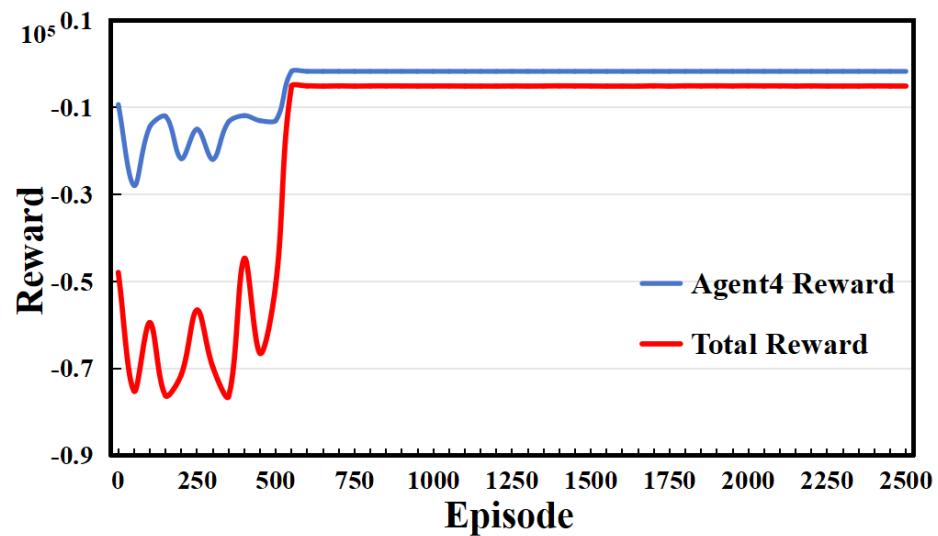


Figure 3. The curve of reward value.

Figure 4 demonstrates the comparisons between DP-MADDPG, MADDPG, and DDPG. As can be observed, DP+MADDPG and MADDPG show good converging performance, while DDPG fails to converge within 2500 episodes. Regarding the converged reward, MADDPG is higher than DP-MADDPG since the introduction of data privacy protection incurs increased operational costs (compromising the solution accuracy). Therefore, the trade-off between privacy protection level and the economy of energy system dispatch is critical.

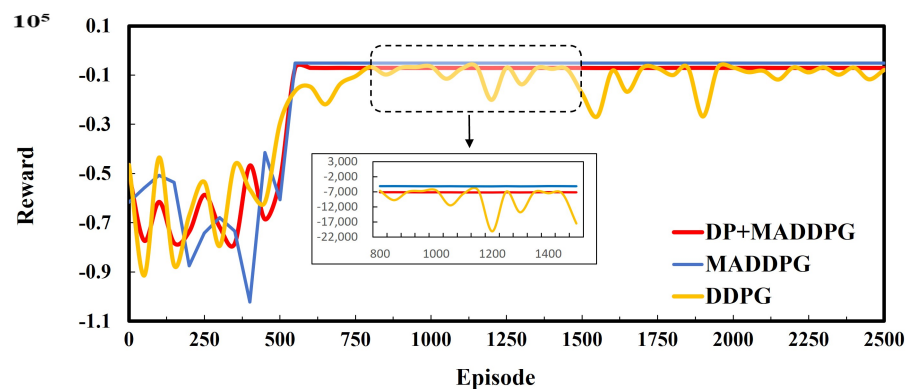


Figure 4. Comparison of different algorithms.

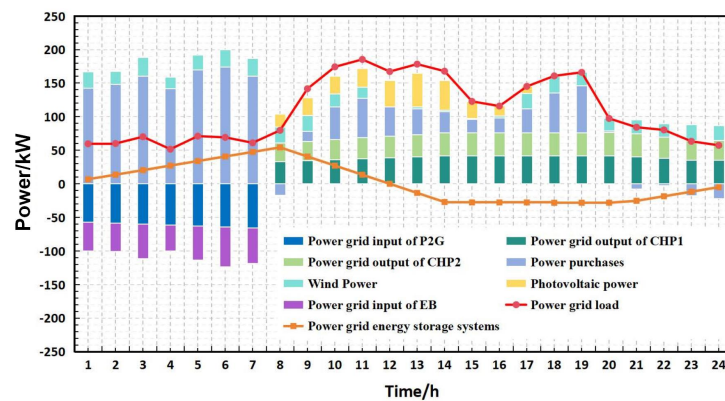
4.2. Optimization Results Analysis

After offline training of the MADDPG algorithm networks using historical data, the trained networks are saved for dynamic economic scheduling of the system. Considering that different intelligent agents have distinct reward evaluation criteria during the training process, this section presents three different energy network models. After completing the training, the power output and exchange power variations in each device within a single period are depicted in the corresponding curves in Figure 5.

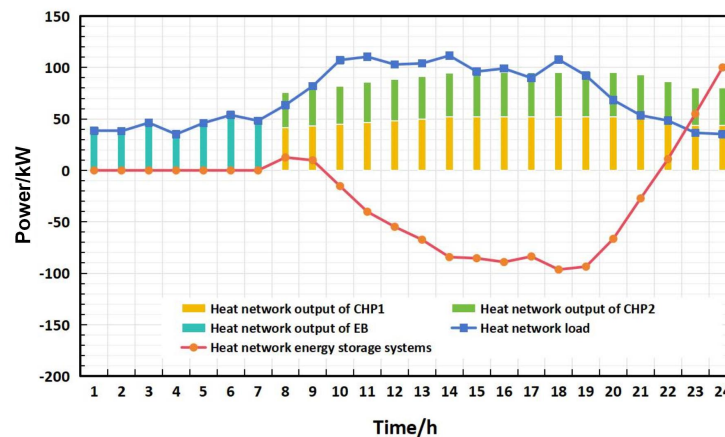
From 0–7 h, the CHP unit is inactive, and the EB device is utilized to provide heat to the heating network, meeting the heat load requirements. The P2G device consumes electricity to supply gas to the gas network, satisfying the gas load and selling excess natural gas for economic benefits. Due to the fluctuation and uncertainty in renewable energy generation, photovoltaic generation is zero during the night, and wind power cannot meet the demands of the grid and other electrical devices. The agent compensates for this power deficit by purchasing electricity from the main grid. Additionally, due to the low electricity prices,

the energy storage system on the grid adopts a strategy of storing electricity to cope with peak power demands, achieving efficient utilization and optimization of energy resources.

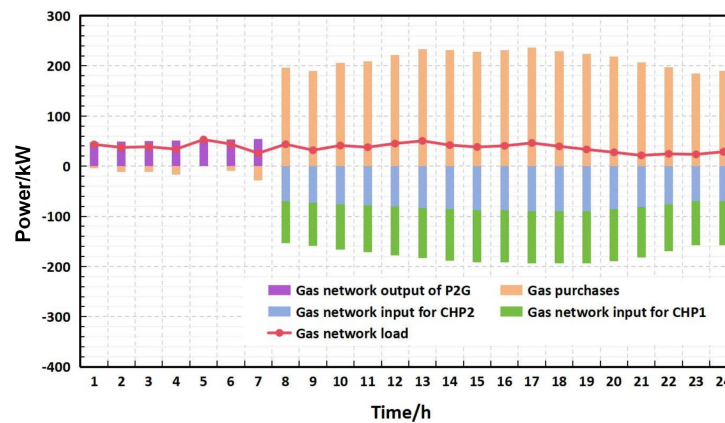
From 8–23 h, the CHP unit starts operating. Due to the high electricity prices, the P2G strategy of converting electricity to gas for the gas network is discontinued, and direct purchase of natural gas is adopted instead. As the electricity and heat demands of the grid and heating network differ, the CHP unit considers the actual conditions of both networks when generating electricity and heat. Therefore, an energy storage system is implemented in the heating network to balance the surplus or deficit of heat. With the involvement of the CHP unit and PV generation, the agent significantly reduces its electricity purchases during periods of high electricity prices compared to the 0–7 h period. Furthermore, the electricity supply curve aligns well with the load demand curve, allowing the energy storage system to operate near its optimal level.



(a) Power network



(b) Heat network



(c) Gas network

Figure 5. The power changes in each network of the EH.

4.3. Privacy Protection Results Analysis

In order to safeguard sensitive information in the power grid from leakage and identification, the privacy data of each round (including gas supply quantities for CHP devices, power-to-gas conversion quantities for P2G, power consumption quantities for EB, load data for the power grid, heating network, and gas network, wind and solar energy generation quantities, and the rate of change of ESS in the power grid and heating network) are protected using differential privacy techniques. The privacy data are perturbed using the Laplace mechanism, and constraints are applied to ensure that the added noise remains within reasonable bounds and does not exceed the power limits of the respective units. The specific transformations are illustrated in Figure 6, which presents the data for 1097 rounds. It can be seen from the figure that, after noise addition, the original data are blurred and distorted, making it impossible to infer and reconstruct specific information.

While maintaining constant gas and electricity consumption, the privacy data are perturbed, and the introduced noise is appropriately translated to the relevant networks based on their coupling relationships. Moreover, an ESS is separately deployed in the power grid, heating network, and gas network, serving as a provider of noise. The variations in energy storage provided by these units are depicted in Figure 7.

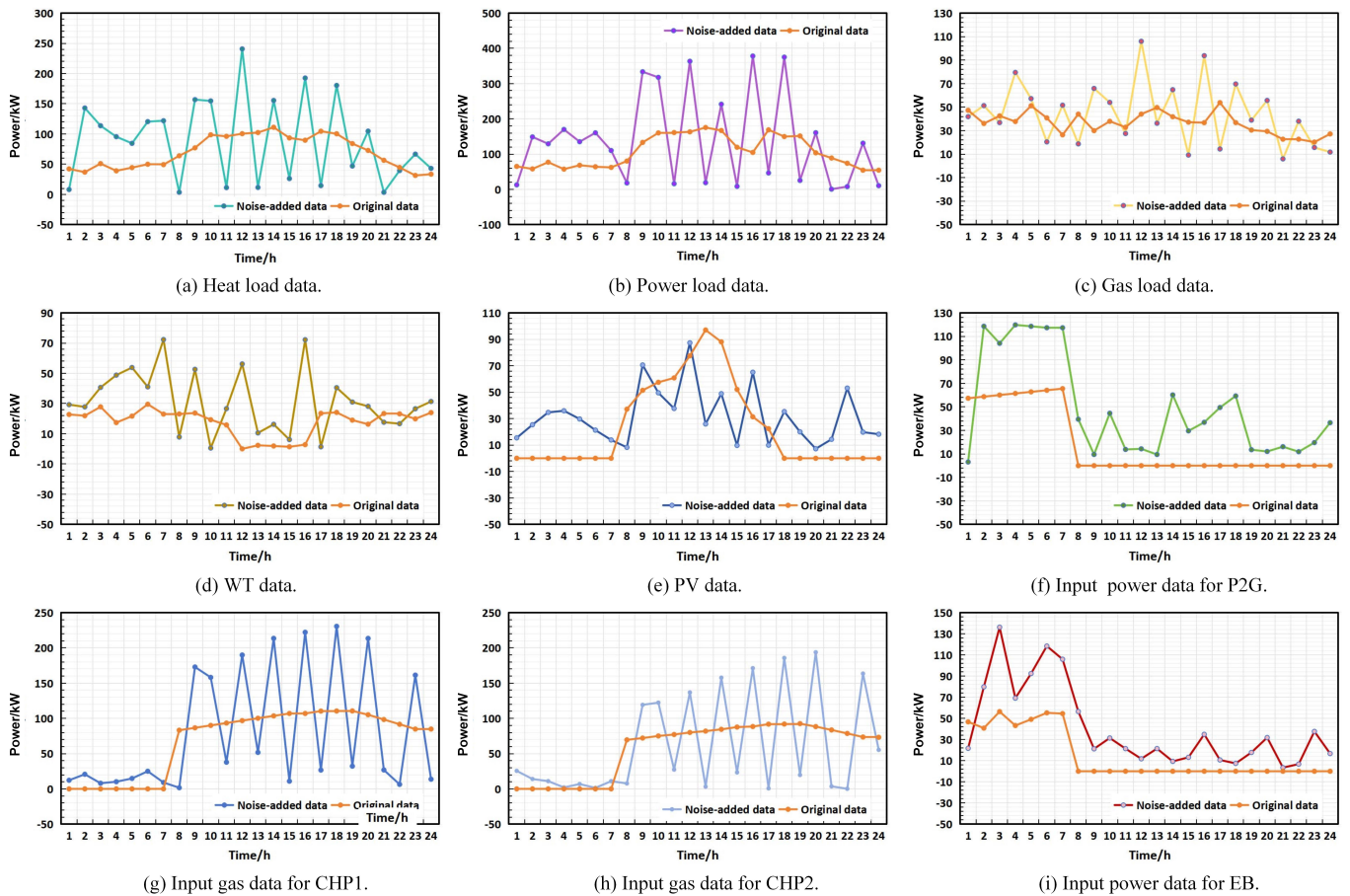


Figure 6. Comparison of original data and synthetic data for a single episode.

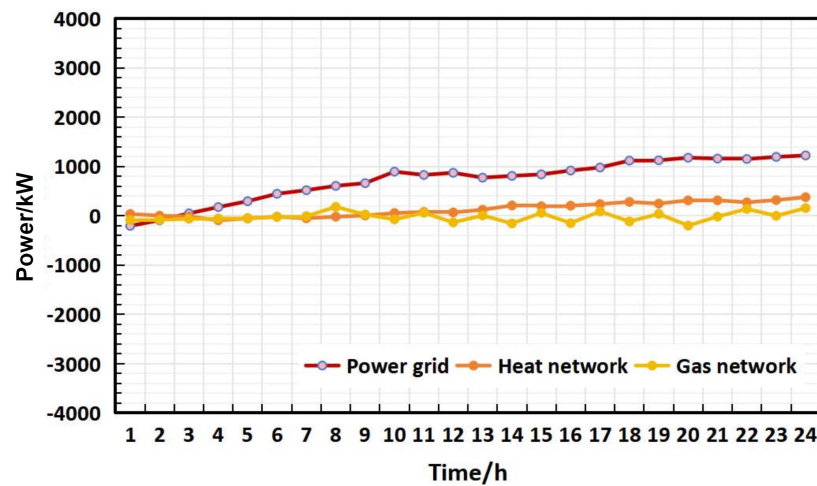


Figure 7. Power changes in noise-canceling energy storage devices.

4.4. Sensitivity Studies

4.4.1. Sensitivities on the Level of Renewable Energy Sources

Considering the stochastic nature of renewable energy generation, it is crucial to assess how well the proposed method handles the inherent uncertainties in renewable energy production. This section provides a comprehensive analysis of the method’s performance under various levels of renewable energy integration. Specifically, four scenarios are selected, including 50% RES, 100% RES, 150% RES, and 200% RES integration. Note that RES data of the previous case studies are used as the benchmark and denoted by 100% RES.

As illustrated in Figure 8, the proposed method shows similar convergence trends for all scenarios, indicating its robustness to the integration of intermittent RES. Based on the results, the converged reward increases with enhancement in RES penetration, which is intuitive since RES is characterized by zero marginal cost, thus reducing the operation costs incurred by fossil fuel consumption.

It should be emphasized that increasing/decreasing the amount of renewable energy generation is equivalent to decreasing/increasing the amount of loads since the net electricity load is equal to actual load minus renewable energy generation. Therefore, we can also interpret from the results that the proposed method shows good robustness in dealing with different levels of loads.

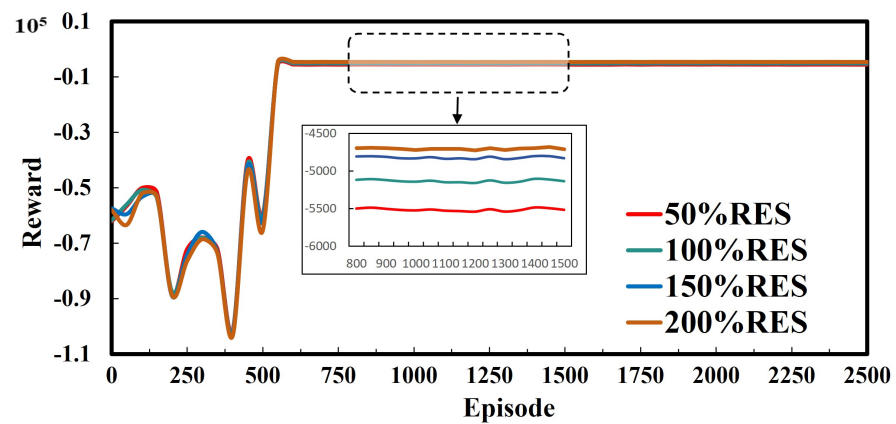


Figure 8. Different levels of renewable energy sources.

4.4.2. Sensitivities on the Number of Agents

In this section, we test four scenarios associated with 1 EH, 2 EHs, 4 EHs, and 8 EHs to observe how many episodes it takes to converge and the computational time. The simu-

lation is performed on a computer with 2-core 3.50 GHz processor and 32 GB RAM, using Python as a tool.

As illustrated in Figure 9, the rewards for all the tested scenarios converge well, where fewer episodes are needed to achieve convergence with the increase in agent amounts. However, the computational time for a single episode increases when more EHs are considered due to the involvement of more variables. Table 4 shows the computational parameters of all the scenarios. Additionally, the converged reward reduces with the increase in EH numbers due to growth in energy consumption of more users.

Table 4. Comparative analysis of different numbers of EHs.

Number of EHs	Computational Time [s]	Average Reward
1	2916	−1830.68
2	4752	−2750.65
4	8424	−5128.89
8	11,232	−9288.73

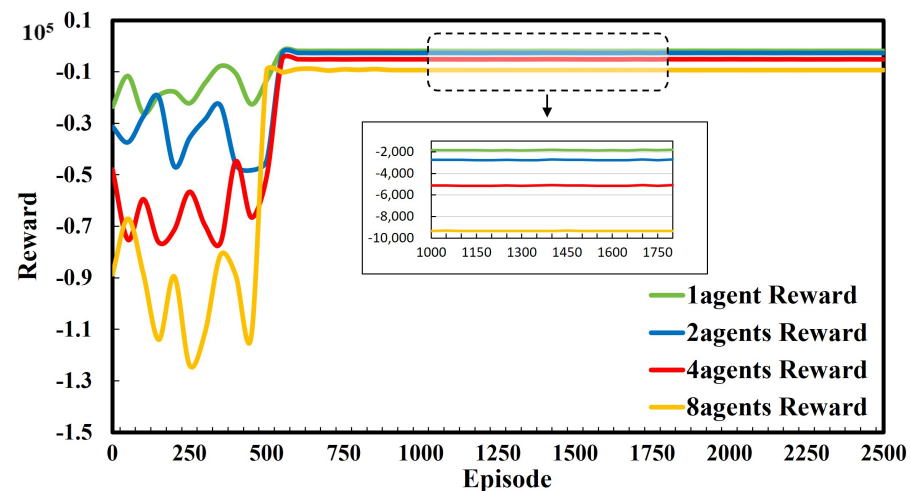


Figure 9. The training process of the proposed algorithm.

Based on these results, it can be concluded that the proposed method can effectively handle the coordination of multiple EBs. It should be emphasized that the MADDPG algorithm has an inherent advantage in solving multi-agent problems; therefore, theoretically, it is capable of handling a problem with many more agents with appropriate model parameter settings.

4.4.3. Sensitivities on Privacy Protection Levels

In this section, we will investigate how different levels of privacy protection impact the computational time and solution accuracy.

In this paper, the parameter epsilon is used to control the degree of noise added to protect data privacy. Increasing the amount of noise added to the original data enhances privacy protection but could lead to notable data distortion and heightened computational overhead. The proposed method is dedicated to securing the privacy of data transmission while concurrently optimizing distributed energy resource scheduling in energy hubs. As demonstrated in Table 5, three key metrics, including discrepancy rate, denoted by σ , computational time, and solution accuracy are used to assess the trade-off between the efficacy of privacy protection and solution performance. High accuracy validates the optimality of the energy dispatch decision made by the algorithm, while a high discrepancy rate ensures the privacy protection efficiency. As can be observed in the table, a larger amount of noises leads to greater discrepancy rates in privacy protection performance,

indicating that more operational costs will be incurred as the consequence of privacy protection. Meanwhile, an increase in computational time is also observed. Therefore, selecting an appropriate epsilon value strikes a balance between privacy preservation and decision accuracy. Determining the suitable amount of noise introduced depends on the specific dataset and privacy requirements.

It should be stressed that, although noise injection reduces the solution accuracy, the physical constraints will not be violated since we introduced the constraint violation penalty, the coefficient of which is particularly large, thus ruling out any solutions located out of the feasibility region of the optimization problem. Additionally, although noise injection causes increased computational complexity, the convergence performance is not essentially impacted.

Table 5. Comparative analysis of different amounts of noise introduced.

Algorithm	Discrepancy Rate σ [%]	Computational Time [s]	Accuracy
MADDPG	0	1685	0.954
	31.2	6242	0.923
DP-MADDPG	63.4	6351	0.902
	90.8	6532	0.883

4.5. Discussion

In the long run, with high-penetration integration of renewable energy driven by carbon neutrality targets, the traditional top-down provision of flexibility from centralized generation units will be insufficient to support efficient accommodation of intermittent and fluctuant wind/solar energy, particularly in the context of low-inertia power systems caused by large-scale displacement of synchronous generators [55,56]. Therefore, it is imperative to exploit the flexibility from DERs as a bottom-up complementary resort [57].

A promising application area for the proposed algorithm is virtual power plants (VPPs), dedicated to arousing the untapped flexibility in DERs by coordinating the operation/response of DERs at the end-user side. VPPs typically involve several hundreds of kW- or MW-magnitude resource aggregation, which can be well-addressed by our algorithm, as illustrated in the case study. Regarding the number of participants (agents), the trained model can effectively handle the coordination of from one to eight energy hubs, as demonstrated, which matches the amount of participants of an average-size VPP in China [58]. Since the training is performed offline, the requirement of hardware and communication delay is not strict. However, due to the involvement of diverse DERs, the control of the VPP may require heterogeneous communication systems and protocols. Additionally, the bottleneck of large-scale VPP application is how to effectively incentivize end users to participate in energy resource aggregation. Additionally, the hierarchical control framework of VPPs highly relies on extensive deployment of sensors for data collection of end users, an efficient algorithm to support the cloud-edge control mode, as well as massive computing power for coordination across numerous end devices. More importantly, incentivizing energy policies are crucial for the participation of end users to provide flexibility to the power system through VPPs [59,60].

At present, the construction of VPPs is primarily restricted at the stage of pilot trials in China [61]. Although there are some commercial VPPs in Western countries running with good profitability, their scales and the amount/diversity of aggregated DERs are limited; therefore, their dependence on a smart and computationally efficient control algorithm to effectively coordinate the operation/response of numerous end users is not urgent. However, with the extensive decommissioning of centralized synchronous gas/coal-fired generation units in recent decades, relative policies are very likely to be mature to fully support the exploitation of flexibilities at the end-user side; VPPs will then play a major role in providing balancing and ancillary services for the power system, a large amount of

heterogeneous DERs will be involved in the control framework of VPPs, and then the proposed control algorithm will truly show its superiority in handling fast scheduling across numerous and diversified resources [62]. However, it is important to emphasize that the proposed algorithm should be regarded as a prototype, the performance of which is tested on the integrated heat–electricity–gas system; however, the physical model/mathematical formulations can be changed to adapt to new policies and emerging technologies without essentially jeopardizing the performance of the algorithm.

5. Conclusions

This paper proposes an EH cluster optimization and scheduling method based on the MADDPG algorithm, targeting EH cluster with multiple IESs. The optimal scheduling problem of the EH cluster is transformed into a deep reinforcement learning model. Each integrated energy system on the EH is treated as an agent, utilizing the capabilities of MADDPG to handle complex tasks with high-dimensional states and action spaces. Through collaborative training among multiple agents within the EH cluster, the method learns cooperative strategies to maximize the performance and efficient utilization of the overall energy system. Additionally, a differential privacy mechanism is introduced in the model to protect sensitive privacy data during the optimization and scheduling process. In each of the three energy networks, a storage system is introduced to serve as a provider of noise, ensuring that the purchase quantities of gas and electricity in the integrated energy system remain unaffected by the introduced noise. Finally, the proposed optimization and scheduling model is applied to a cluster scheduling optimization problem consisting of four EH. The experimental simulations demonstrate that the proposed method offers reasonable optimization strategies for scheduling problems and exhibits good generalization capabilities when facing uncertain fluctuations in renewable energy output.

In the future, we will explore multi-level optimization for the hierarchical control structure of VPPs, where multiple lower-level agents interact with the higher-level controller.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z. and Q.W.; software, X.Z. and Q.W.; validation, J.Y. and Q.S.; formal analysis, X.Z. and Q.W.; investigation, Q.S. and X.L.; resources, Q.W.; data curation, Q.W.; writing—original draft preparation, X.Z. and Q.W.; writing—review and editing, J.Y., Q.S. and H.H.; visualization, X.Z. and Q.W.; supervision, X.Z.; project administration, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Program of State Grid “Research of Interactive Control between Distributed Energy Resources and Mega-City Grids under Multi-Constraints”. The grant number is 5700-202311602A-3-2-ZN.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Xi Zhang, Qinghe Sun and Heng Hu are full-time employees of State Grid Smart Grid Research Institute Co., Ltd. Qiong Wang is a full-time employee of State Grid Beijing Municipal Electric Power Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DDPG	Deep deterministic policy gradient
DRL	Deep reinforcement learning
DP	Differential privacy
EH	Energy hub
ESS	Energy storage system
HE	Homomorphic encryption
IES	Integrated energy system
MADDPG	Multi-agent deep deterministic policy gradient

MADRL	Multi-agent deep reinforcement learning
MDP	Markov decision process
MG	Microgrid
MILP	Mixed-integer linear programming
RL	Reinforcement learning
Indices and Sets	
n, \mathcal{N}	Index/set of CHP from 1 to N .
i, \mathcal{M}	Index/set of EH nodes from 1 to M .
t, \mathcal{T}	Index/set of time slots from 1 to T .
S_i	Set of the state for agent i from time 1 to t , i.e., $\{s_{i,1}, s_{i,2}, \dots, s_{i,t}\}$.
A_i	Set of the action for agent i from time 1 to t , i.e., $\{a_{i,1}, a_{i,2}, \dots, a_{i,t}\}$.
$\alpha_e^{ES}, \alpha_h^{ES}, \alpha_g^{ES}$	The self-discharge efficiency of electricity, heat, and gas networks.
$\alpha_e^{ES}, \alpha_h^{ES}, \alpha_g^{ES}$	The self-discharge efficiency of electricity, heat, and gas networks.
$\eta_e^{ch}, \eta_h^{ch}, \eta_g^{ch}$	The charge efficiency of electricity, heat, and gas networks.
$\eta_e^{dis}, \eta_h^{dis}, \eta_g^{dis}$	The discharge efficiency of electricity, heat, and gas networks.
η_e^{CHP}	The the gas-to-electricity conversion efficiencies.
Parameters	
η_h^{CHP}	The the gas-to-heat conversion efficiencies.
η^{EB}	The the power-to-heat conversion efficiencies.
η^{P2G}	The the power-to-gas conversion efficiencies.
$\pi_t^{CO_2}$	The price of carbon trading.
C^{CHP}	The O&M costs for CHP.
C^{EB}	The O&M costs for EB.
C^{ES}	The O&M costs for ES.
C^{P2G}	The O&M costs for P2G.
π_t^e, π_t^g	The price of electricity and gas at time t .
γ, λ	The small/large positive value as a reward weight.
e^{CHP}	The carbon emission quota associated with the unit energy generated.
E^{CHP}	The carbon emission intensity associated with the unit energy generated.
E_{min}^x, E_{max}^x	The lower/upper limits of the ESS's power.
$G_{min}^{CHP}, G_{max}^{CHP}$	The lower/upper bounds of gas consumption.
G_{ramp}^{CHP}	The maximum ramping power of the n th CHP.
P_{ramp}^{CHP}	The maximum ramping power of the P2G.
$P_{min}^{EB}, P_{max}^{EB}$	The lower/upper limits of the EB's power.
P_{ramp}^{EB}	The maximum ramping power of the EB.
$P_{min}^{P2G}, P_{max}^{P2G}$	The lower/upper limits of the P2G's power.
$P_{min}^{x,ch}, P_{max}^{x,dis}$	The lower/upper limits of charging/discharging power.
Variables	
$E_{i,t}^x, E_{x,t}^{ES}$	The stored energy of ESS or networks at time t .
$H_{n,i,t}^{CHP}$	The heat output of the n th CHP in node i at time t .
$H_{i,t}^{EB}$	The thermal output of EB in node i .
$G_{n,i,t}^{CHP}$	The gas consumption of the n th CHP in node i at time t .
$G_{i,t}^{P2G}$	The gas output of P2G in node i .
$\Delta G_{n,i,t}^{CHP}$	The variation of $G_{n,i,t}^{CHP}$ from slot t to $t + \Delta t$.
$G_{i,t}^{net}$	The exchanged power with the external gas network in node i at time t .
$L_{i,t}^e, L_{i,t}^h, L_{i,t}^g$	The values of electrical, thermal, and gas load at time t .
$P_{n,i,t}^{CHP}$	The power output of the n th CHP in node i at time t .
$P_{i,t}^{EB}$	The electric power input of EB in node i .
$\Delta P_{i,t}^{EB}$	The variation of $P_{i,t}^{EB}$ from slot t to $t + \Delta t$.
$P_{i,t}^{P2G}$	The electric power input of P2G in node i .
$\Delta P_{i,t}^{P2G}$	The variation of $P_{i,t}^{P2G}$ from slot t to $t + \Delta t$.
$P_{i,t}^{PV}$	The power generation of the photoelectric power generation unit in node i at time t .
$P_{i,t}^{WT}$	The power generation of the wind power generation unit in node i at time t .
$P_{i,t}^{x,ch}, P_{i,t}^{x,dis}$	The charging/discharging power in node i at time t .
$P_{i,t}^{net}$	The exchanged power with the main grid in node i at time t .

References

1. Wang, Y.; Hu, J.; Liu, N. Energy Management in Integrated Energy System Using Energy–Carbon Integrated Pricing Method. *IEEE Trans. Sustain. Energy* **2023**, *14*, 1992–2005. [[CrossRef](#)]
2. Liu, N.; Tan, L.; Sun, H.; Zhou, Z.; Guo, B. Bilevel Heat–Electricity Energy Sharing for Integrated Energy Systems With Energy Hubs and Prosumers. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3754–3765. [[CrossRef](#)]
3. Yan, C.; Bie, Z.; Liu, S.; Urgun, D.; Singh, C.; Xie, L. A reliability model for integrated energy system considering multi-energy correlation. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 811–825. [[CrossRef](#)]
4. Zhang, Z.; Wang, C.; Yang, M.; Chen, X.; Lv, H. Day-ahead Optimal Dispatch for Integrated Energy System Considering Power-to-gas and Dynamic Pipeline Networks. *IEEE Ind. Appl. Soc. Annu. Meet.* **2021**, *57*, 3317–3328. [[CrossRef](#)]
5. Alabdulwahab, A.; Abusorrah, A.; Zhang, X.; Shahidehpour, M. Coordination of interdependent natural gas and electricity infrastructures for firming the variability of wind energy in stochastic day-ahead scheduling. *IEEE Trans. Sustain. Energy* **2015**, *6*, 606–615. [[CrossRef](#)]
6. Quelhas, A.; Gil, E.; McCalley, J.D.; Ryan, S.M. A multiperiod generalized network flow model of the US integrated energy system: Part I—Model description. *IEEE Trans. Power Syst.* **2007**, *22*, 829–836. [[CrossRef](#)]
7. Rinaldi, S.M.; Peerenboom, J.P.; Kelly, T.K. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Syst. Mag.* **2001**, *21*, 11–25.
8. Shahidehpour, M.; Fu, Y.; Wiedman, T. Impact of natural gas infrastructure on electric power systems. *Proc. IEEE* **2005**, *93*, 1042–1056. [[CrossRef](#)]
9. Du, X.; Wu, Z.; Zou, L.; Tang, Y.; Fang, C.; Wang, C. Optimal Configuration of Integrated Energy Systems Based on Mixed Integer Linear Programming. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2021; Volume 2, pp. 242–246.
10. Laraki, M.H.; Brahmi, B.; El-Bayeh, C.Z.; Rahman, M.H. Energy management system for a Stand-alone Wind/Diesel/BESS/Fuel-cell Using Dynamic Programming. In Proceedings of the 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), Monastir, Tunisia, 22–25 March 2021; pp. 1258–1263.
11. Zheng, J.; Wu, Q.; Jing, Z. Coordinated scheduling strategy to optimize conflicting benefits for daily operation of integrated electricity and gas networks. *Appl. Energy* **2017**, *192*, 370–381. [[CrossRef](#)]
12. Lei, Y.; Xu, J.; Tong, N.; Shen, M. An Economically Optimized Planning of Regional Integrated Energy System Considering Renewable Energy and Energy Storage. In Proceedings of the 2022 IEEE PES 14th Asia-Pacific Power and Energy Engineering, Melbourne, Australia, 22–23 November 2022; pp. 1–6.
13. Li, C.; Yang, H.; Shahidehpour, M.; Xu, Z.; Zhou, B.; Cao, Y.; Zeng, L. Optimal Planning of Islanded Integrated Energy System With Solar-Biogas Energy Supply. *IEEE Trans. Sustain. Energy* **2020**, *11*, 2437–2448. [[CrossRef](#)]
14. Liu, X.; Wu, J.; Jenkins, N.; Bagdanavicius, A. Combined analysis of electricity and heat networks. *Appl. Energy* **2016**, *162*, 1238–1250. [[CrossRef](#)]
15. Shi, M.; Wang, H.; Xie, P.; Lyu, C.; Jian, L.; Jia, Y. Distributed Energy Scheduling for Integrated Energy System Clusters With Peer-to-Peer Energy Transaction. *IEEE Trans. Smart Grid* **2023**, *14*, 142–156. [[CrossRef](#)]
16. Zhang, Y.; Mou, Z.; Gao, F.; Jiang, J.; Ding, R.; Han, Z. UAV-enabled secure communications by multi-agent deep reinforcement learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 11599–11611. [[CrossRef](#)]
17. Tang, F.; Kawamoto, Y.; Kato, N.; Liu, J. Future intelligent and secure vehicular network toward 6G: Machine-learning approaches. *Proc. IEEE* **2019**, *108*, 292–307. [[CrossRef](#)]
18. Kha, Q.H.; Ho, Q.T.; Le, N.Q.K. Identifying SNARE proteins using an alignment-free method based on multiscan convolutional neural network and PSSM profiles. *J. Chem. Inf. Model.* **2022**, *62*, 4820–4826. [[CrossRef](#)]
19. Le, N.Q.K. Potential of deep representative learning features to interpret the sequence information in proteomics. *Proteomics* **2022**, *22*, 2100232. [[CrossRef](#)]
20. Li, S.; Wu, Y.; Cui, X.; Dong, H.; Fang, F.; Russell, S. Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 4213–4220.
21. Lei, W.; Wen, H.; Wu, J.; Hou, W. MADDPG-based security situational awareness for smart grid with intelligent edge. *Appl. Sci.* **2021**, *11*, 3101. [[CrossRef](#)]
22. Qiu, D.; Dong, Z.; Zhang, X.; Wang, Y.; Strbac, G. Safe reinforcement learning for real-time automatic control in a smart energy-hub. *Appl. Energy* **2022**, *309*, 118403. [[CrossRef](#)]
23. Zhang, T.; Sun, M.; Qiu, D.; Zhang, X.; Strbac, G.; Kang, C. A Bayesian Deep Reinforcement Learning-based Resilient Control for Multi-Energy Micro-grid. *IEEE Trans. Power Syst.* **2023**, *38*, 5057–5072. [[CrossRef](#)]
24. Xu, Z.; Han, G.; Liu, L.; Martínez-García, M.; Wang, Z. Multi-Energy Scheduling of an Industrial Integrated Energy System by Reinforcement Learning-Based Differential Evolution. *IEEE Trans. Green Commun. Netw.* **2021**, *5*, 1077–1090. [[CrossRef](#)]
25. Park, L.; Lee, C.; Kim, J.; Mohaisen, A.; Cho, S. Two-stage IoT device scheduling with dynamic programming for energy Internet systems. *IEEE Internet Things J.* **2019**, *6*, 8782–8791. [[CrossRef](#)]
26. Zhou, Y.; Jia, L.; Zhao, Y.; Zhan, Z. Optimal dispatch of an integrated energy system based on deep reinforcement learning considering new energy uncertainty. *Energy Rep.* **2023**, *4*, 804–809.

27. Wang, Y.; Yang, Z.; Dong, L.; Huang, S.; Zhou, W. Energy Management of Integrated Energy System Based on Stackelberg Game and Deep Reinforcement Learning. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–1 November 2020; Volume 2, pp. 2645–2651.
28. Wang, X.; Chen, S.; Yan, D.; Wei, J.; Yang, Z. Multi-agent deep reinforcement learning-based approach for optimization in microgrid clusters with renewable energy. In Proceedings of the 2021 International Conference on Power System Technology (POWERCON), Haikou, China, 8–9 December 2021; Volume 4, pp. 413–419.
29. Ren, H.; Gao, W. A MILP model for integrated plan and evaluation of distributed energy systems. *Appl. Energy* **2010**, *87*, 1001–1014. [[CrossRef](#)]
30. Nebuloni, R.; Meraldi, L.; Bovo, C.; Ilea, V.; Berizzi, A.; Sinha, S.; Tamirisakandala, R.B.; Raboni, P. A hierarchical two-level MILP optimization model for the management of grid-connected BESS considering accurate physical model. *Appl. Energy* **2023**, *334*, 120697. [[CrossRef](#)]
31. Chen, Y.; Han, W.; Zhu, Q.; Liu, Y.; Zhao, J. Target-driven obstacle avoidance algorithm based on DDPG for connected autonomous vehicles. *EURASIP J. Adv. Signal Process.* **2022**, *2022*, 61. [[CrossRef](#)]
32. Fan, P.; Ke, S.; Yang, J.; Li, R.; Li, Y.; Yang, S.; Liang, J.; Fan, H.; Li, T. A load frequency coordinated control strategy for multimicrogrids with V2G based on improved MA-DDPG. *Int. J. Electr. Power Energy Syst.* **2023**, *146*, 108765. [[CrossRef](#)]
33. Ao, T.; Zhang, K.; Shi, H.; Jin, Z.; Zhou, Y.; Liu, F. Energy-Efficient Multi-UAVs Cooperative Trajectory Optimization for Communication Coverage: An MADRL Approach. *Remote Sens.* **2023**, *15*, 429. [[CrossRef](#)]
34. Wang, S.; Duan, J.; Shi, D.; Xu, C.; Li, H.; Diao, R.; Wang, Z. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. *IEEE Trans. Power Syst.* **2020**, *35*, 4644–4654. [[CrossRef](#)]
35. Peng, H.; Shen, X. Multi-agent reinforcement learning based resource management in MEC-and UAV-assisted vehicular networks. *IEEE J. Sel. Areas Commun.* **2020**, *39*, 131–141. [[CrossRef](#)]
36. Lee, H.; Jeong, J. Multi-agent deep reinforcement learning (MADRL) meets multi-user MIMO systems. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; Volume 4, pp. 1–6.
37. Kóvári, B.; Lovétei, I.; Aradi, S.; Bécsi, T. Multi-Agent Deep Reinforcement Learning (MADRL) for Solving Real-Time Railway Rescheduling Problem. In Proceedings of the the Fifth International Conference on Railway Technology, Montpellier, France, 22–25 August 2022; Volume 1, pp. 1–6.
38. Zhang, C.; Ahmad, M.; Wang, Y. ADMM based privacy-preserving decentralized optimization. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 565–580. [[CrossRef](#)]
39. Yuan, Z.P.; Li, P.; Li, Z.L.; Xia, J. A Fully Distributed Privacy-Preserving Energy Management System for Networked Microgrid Cluster Based on Homomorphic Encryption. *IEEE Trans. Smart Grid* **2023**, *2*, 1. [[CrossRef](#)]
40. Nozari, E.; Tallapragada, P.; Cortés, J. Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design. *Automatica* **2017**, *81*, 221–231. [[CrossRef](#)]
41. Cheng, Z.; Ye, F.; Cao, X.; Chow, M.Y. A Homomorphic Encryption-Based Private Collaborative Distributed Energy Management System. *IEEE Trans. Smart Grid* **2021**, *12*, 5233–5243. [[CrossRef](#)]
42. Zhang, T.; Zhu, T.; Xiong, P.; Huo, H.; Tari, Z.; Zhou, W. Correlated Differential Privacy: Feature Selection in Machine Learning. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2115–2124. [[CrossRef](#)]
43. Zhu, T.; Ye, D.; Wang, W.; Zhou, W.; Yu, P.S. More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 2824–2843. [[CrossRef](#)]
44. Aziz, R.; Banerjee, S.; Bouzefrane, S.; Le Vinh, T. Exploring Homomorphic Encryption and Differential Privacy Techniques towards Secure Federated Learning Paradigm. *Future Internet* **2023**, *15*, 310. [[CrossRef](#)]
45. He, T.; Wu, X.; Dong, H.; Guo, F.; Yu, W. Distributed Optimal Power Scheduling for Microgrid System via Deep Reinforcement Learning with Privacy Preserving. In Proceedings of the 2022 IEEE 17th International Conference on Control & Automation (ICCA), Naples, Italy, 27–30 June 2022; Volume 1, pp. 820–825.
46. Wang, Z.; Wan, R.; Gui, X.; Zhou, G. Deep Reinforcement Learning of Cooperative Control with Four Robotic Agents by MADDPG. In Proceedings of the 2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC), Chongqing, China, 6–8 November 2020; Volume 5, pp. 287–290.
47. Mao, H.; Zhang, Z.; Xiao, Z.; Gong, Z. Modelling the dynamic joint policy of teammates with attention multi-agent DDPG. *arXiv* **2018**, arXiv:1811.07029.
48. Li, Y.; Wang, B.; Yang, Z.; Li, J.; Chen, C. Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via Stackelberg game. *Appl. Energy* **2022**, *308*, 118392. [[CrossRef](#)]
49. Yang, M.; Tjuawinata, I.; Lam, K.Y. K-Means Clustering With Local ϵ -Privacy for Privacy-Preserving Data Analysis. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 2524–2537. [[CrossRef](#)]
50. Yang, W.; Lam, K.Y. Automated cyber threat intelligence reports classification for early warning of cyber attacks in next generation SOC. In Proceedings of the International Conference on Information and Communications Security 2019, Beijing, China, 15–17 December 2020; Volume 2, pp. 145–164.
51. Lakshmi, R.; Baskar, S. Efficient text document clustering with new similarity measures. *Int. J. Bus. Intell. Data Min.* **2021**, *18*, 49–72. [[CrossRef](#)]

52. Wang, S.; Zhang, X.; Cheng, Y.; Jiang, F.; Yu, W.; Peng, J. A fast content-based spam filtering algorithm with fuzzy-SVM and K-means. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–17 January 2018; Volume 14, pp. 301–307.
53. Ghezelbash, R.; Maghsoudi, A.; Carranza, E.J.M. Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm. *Comput. Geosci.* **2020**, *134*, 104335. [[CrossRef](#)]
54. Pradana, M.G.; Ha, H.T. Maximizing strategy improvement in mall customer segmentation using k-means clustering. *J. Appl. Data Sci.* **2021**, *2*, 19–25. [[CrossRef](#)]
55. Du, W.; Bi, J.; Wang, T.; Wang, H. Impact of grid connection of large-scale wind farms on power system small-signal angular stability. *CSEE J. Power Energy Syst.* **2015**, *1*, 83–89. [[CrossRef](#)]
56. Munkhchuluun, E.; Meegahapola, L.; Vahidnia, A. Long-term voltage stability with large-scale solar-photovoltaic (PV) generation. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105663. [[CrossRef](#)]
57. Mondal, A.; Illindala, M.S. Improved frequency regulation in an islanded mixed source microgrid through coordinated operation of DERs and smart loads. *IEEE Trans. Ind. Appl.* **2017**, *54*, 112–120. [[CrossRef](#)]
58. Li, Y.; Gao, W.; Ruan, Y. Feasibility of virtual power plants (VPPs) and its efficiency assessment through benefiting both the supply and demand sides in Chongming country, China. *Sustain. Cities Soc.* **2017**, *35*, 544–551. [[CrossRef](#)]
59. Rouzbahani, H.M.; Karimipour, H.; Lei, L. A review on virtual power plant for energy management. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101370. [[CrossRef](#)]
60. Zamani, A.G.; Zakariazadeh, A.; Jadid, S.; Kazemi, A. Stochastic operational scheduling of distributed energy resources in a large scale virtual power plant. *Int. J. Electr. Power Energy Syst.* **2016**, *82*, 608–620. [[CrossRef](#)]
61. Tan, C.; Tan, Z.; Wang, G.; Du, Y.; Pu, L.; Zhang, R. Business model of virtual power plant considering uncertainty and different levels of market maturity. *J. Clean. Prod.* **2022**, *362*, 131433. [[CrossRef](#)]
62. Shabanzadeh, M.; Sheikh-El-Eslami, M.K.; Haghifam, M.R. A medium-term coalition-forming model of heterogeneous DERs for a commercial virtual power plant. *Appl. Energy* **2016**, *169*, 663–681. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.