

Article

Secondary Voltage Collaborative Control of Distributed Energy System via Multi-Agent Reinforcement Learning

Tianhao Wang ¹, Shiqian Ma ¹, Na Xu ^{2,*}, Tianchun Xiang ³, Xiaoyun Han ², Chaoxu Mu ² and Yao Jin ³

¹ Electric Power Research Institute, State Grid Tianjin Electric Power Company, No. 8, Haitai Huake 4th Road, Huayuan Industrial Zone, Binhai High Tech Zone, Tianjin 300384, China

² Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

³ State Grid Tianjin Electric Power Company, No. 39 Wujing, Guangfu Street, Hebei District, Tianjin 300010, China

* Correspondence: xuna1205@tju.edu.cn

Abstract: In this paper, a new voltage cooperative control strategy for a distributed power generation system is proposed based on the multi-agent advantage actor-critic (MA2C) algorithm, which realizes flexible management and effective control of distributed energy. The attentional actor-critic message processor (AACMP) is extended into the MA2C method to select the important messages from all communication messages adaptively and process important messages efficiently. The cooperative control strategy trained by centralized training and decentralized execution frame will take over the responsibility of the secondary control level for voltage restoration in a distributed manner. The introduction of the attention mechanism reduces the amount of information exchanged and the requirements of the communication network. Finally, a distributed system with six energy nodes is used to verify the effectiveness of the proposed control strategy.

Keywords: distributed energy; deep reinforcement learning; attentional mechanism; nodal voltage; coordination optimization



Citation: Wang, T.; Ma, S.; Xu, N.; Xiang, T.; Han, X.; Mu, C.; Jin, Y. Secondary Voltage Collaborative Control of Distributed Energy System via Multi-Agent Reinforcement Learning. *Energies* **2022**, *15*, 7047. <https://doi.org/10.3390/en15197047>

Academic Editor: Abu-Siada Ahmed

Received: 27 August 2022

Accepted: 15 September 2022

Published: 25 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

With the development of economy and technology, especially the active promotion and application of renewable energy, the energy demand of users has begun to diversify [1]. The development of renewable energy has penetrated into all walks of life, especially in the power industry and energy-efficient buildings [2]. In order to adapt to the new market demand, there are higher requirements for the development and efficient utilization of renewable energy in many countries. At the same time, with the development and maturity of various electricity technologies, more and more technologies can be selected. The hierarchical optimization method has been proposed to realize the complementary nature of different energies [3,4]. The optimal dispatching model with battery energy storage was proposed based on mixed integer programming [5,6]. As a type of open energy system, distributed energy systems have begun to display a multi-functional trend, which not only contains a variety of energy inputs but can also satisfy a variety of energy needs of users at the same time [7–9]. The distributed energy system is relative to the traditional centralized energy supply system. The traditional centralized energy supply adopts high-capacity equipment and centralized production. It needs to transmit various energy to a large range of users through special transmission facilities (large power grid, large heat grid, etc.). The distributed energy system is a small and medium-sized energy conversion and utilization system that directly faces users, which can produce and supply energy locally according to users' needs by designing multiple functions to satisfy multiple objectives [10,11]. The rapid development of various new energy sources has led users to constantly put forward new requirements for energy systems, mainly including high efficiency, reliability, economy, environmental protection, sustainable development, etc. [12,13]. The new distributed

energy system can better meet these requirements at the same time and achieve multiple functional objectives by selecting appropriate technologies through system optimization and integration. The rapid development of renewable energy and the deep integration of information and intelligent technologies present a new trend of integrating information and energy and integrating intelligence and materials [14]. Due to the randomness and volatility of renewable energy sources, the stability of voltage is very important for the distributed energy sources system. Currently, there are three control strategies for inverters, which are as follows: constant voltage constant frequency (V/F) control and constant power (PQ) control, and sagging control (droop control) [15]. The droop control simulations of synchronous generator characteristics rapidly gain the stability of voltage and current to “plug and play”. In order to satisfy the different dynamic characteristics of distributed energy, a generalized droop control was applied to promote the rapidity and accuracy of tracking voltage and frequency reference values [16]. Existing literature points out that the multi-input, multi-output control strategy can improve the stability of a distributed energy resources unit [17]. The structure of a distributed energy system with multiple renewable energy sources is similar to that of a multi-agent system [18,19]. Thus, multi-agent system technology has been used to implement load frequency control, and the dynamic performance of the system was effectively improved [20].

In a distributed energy system, there are various coupling modes between different forms of energy, and the energy supply combinations and operation strategies are different. How to design the coordination mechanism of a distributed energy system is an important prerequisite for the collaborative optimization and operation of a distributed energy system [21]. In [22], an intelligent control strategy based on reinforcement learning was proposed to realize the reuse of renewable energy and suppress the frequency fluctuation caused by uncertainty. Each distributed energy node is considered as an agent, and each agent has certain independence and autonomy. A multi-agent-based system can realize the multi-energy cooperative operation of a distributed energy system [23]. Applying the multi-agent reinforcement learning algorithm to the distributed energy system is a research hotspot in the field of control [24]. From this, a frequency control method based on a distributed multi-agent is proposed. Each agent can communicate with its neighbour agents based on power line carrier communication technology. The control objective of a cooperative frequency recovery is realized by adopting the optimized average consensus algorithm [25–27]. Considering the inherent trade-off between power sharing and voltage and frequency regulation caused by traditional droop control methods, automatic generation control and automatic voltage control have been proposed based on a multi-agent system to achieve cooperative control of active and reactive power [28–31]. Multi-agent deep reinforcement learning has been effectively applied in many fields [32–35]. However, due to the high permeability of renewable energy in distributed information energy systems, the basic architecture of physical information models and effective communication methods in distributed information energy system still need to be analyzed [36,37].

In this paper, the exposition is launched in the following structure. The multi-agent model construction method of distributed energy system is introduced in Section 2. In Section 3, we apply a multi-agent actor-critic algorithm to the secondary voltage control of a distributed energy system. In order to reduce communication pressure and improve the efficiency of processing information from neighboring agents, we add an attentional actor-critic message processor [38]. The simulation results are shown in Section 4. Finally, we conclude with the main thesis of the research and development work of summing up and the further research work in the paper.

2. Distributed Energy System

The distributed energy system is located near the users and directly faces the needs of users. It can effectively simplify the transmission link of the system to provide users with energy, thus reducing the energy loss and transmission cost in the process of energy transmission and increasing the security of users' energy supply. As a new energy supply

mode, a distributed energy system is a powerful supplement to a centralized power supply system. The energy node is the physical node in the energy flow level, and the agent is the virtual agent at the information level. A model of a distributed energy system is shown in Figure 1.

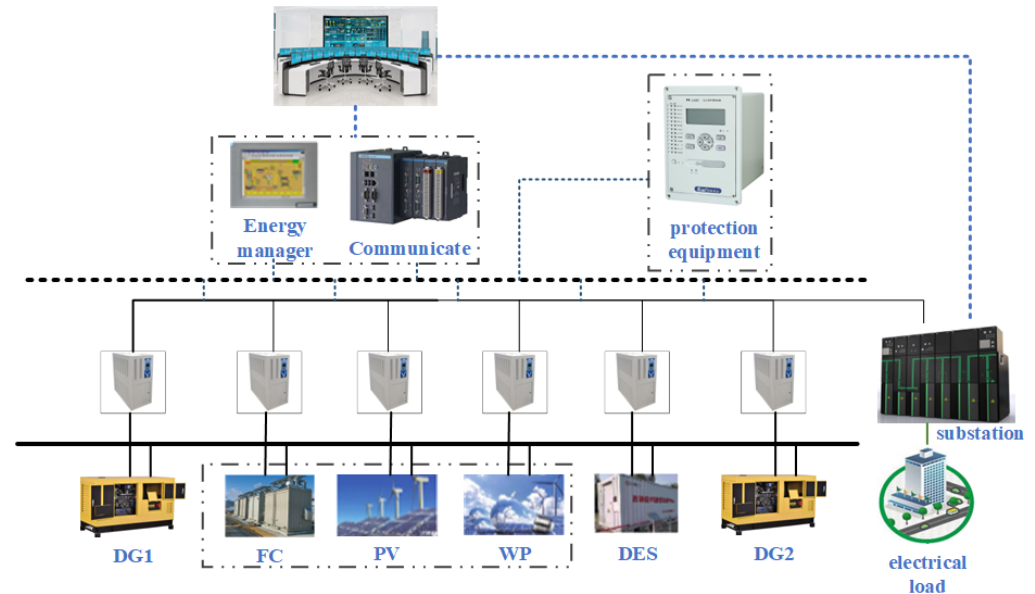


Figure 1. Distributed energy system model.

In this paper, we adopt a simple structure for the distributed energy system. It contains a central management center, power generation equipment, electrical load and power electronics. The system is composed of a multi-energy coordinated and complementary energy network, with solar, wind and other renewable energy generation as the primary source and thermal power generation as an auxiliary source. Renewable energy is incorporated into the power grid through the electronic power conversion device, and the inverter is used to convert the electric energy. The central management system is designed to provide real-time information for operation scheduling, carry out scheduling decision-making and realize autonomous management of the distributed energy system.

The control structure of the distributed energy system is shown in Figure 2. The control structure considers each energy node as an agent in a multi-agent system. The secondary voltage control based on the voltage source inverter depends on agents cooperating with each other rather than relying on the centralized controller. In addition to local information, each agent can obtain information from neighboring agents through the communication mechanism. The set points of the primary control of each energy node are obtained by its own distributed secondary controller. The primary control requires each energy node to automatically distribute active and reactive loads using local information while sustaining voltage and frequency stability, which can generally be achieved by $p - \omega$ and $Q - V$ droop control methods.

$$\omega = \omega_0 - D_p P \quad (1)$$

$$v = V_0 - D_q Q \quad (2)$$

where ω is the frequency generated based on droop control method. ω_0 and V_0 are the set point of the primary control, which is determined by the secondary control. D_p and D_q are the corresponding droop coefficients, which are selected based on the active and reactive power rating generated by the coordination of each distributed generator. Considering the randomness and model uncertainty of renewable energy, the multi-agent reinforcement learning method is introduced to achieve the secondary voltage cooperative control strategy of each agent so the node voltage can recover to the reference value. The detailed algorithm design method will be introduced in the next chapter.

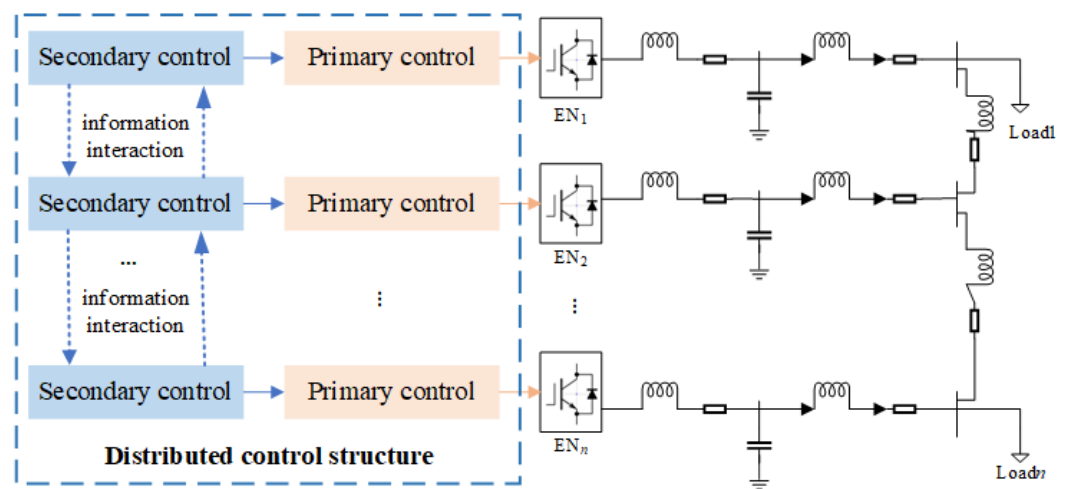


Figure 2. Distributed scheme for secondary voltage control.

3. MA2C-Attention Algorithm

As we all know that the core idea of reinforcement learning is “trial-and-error”, the agent iteratively optimizes based on the feedback it receives through interaction with the environment [28]. In the RL domain, the problem to be resolved is commonly described as a Markov decision process. When multiple agents interact with the environment at the same time, the whole system becomes a multi-agent system, which is able to resolve complex tasks through the collaboration of individual agents. Each agent still follows the goal of reinforcement learning, which is to maximize the cumulative returns that can be obtained, and the change in the global state of the environment is related to the joint actions of all agents. Therefore, the influence of joint action should be considered in the process of agent strategy learning. In the distributed information energy system, the energy node is comparable to an agent with perceptive ability. The cost function and strategy can be optimized by using the multi-agent reinforcement learning method and iterative learning of the reward function.

The distributed energy system is constructed as a multi-agent network, where each energy node is controlled by a local RL agent. The state space of each energy node is designed as:

$$s_{i,t} = (P_i, Q_i, \delta_i, i_{odi}, i_{oqi}, i_{bdi}, i_{bqi}, v_{bdi}, v_{bqi}) \quad (3)$$

where P_i and Q_i are the active power and reactive power of the i th distributed generation unit, respectively. δ_i is the reference frame with respect to the common reference frame. i_{odi} and i_{oqi} are the direct and quadrature output current components of the i th distributed generation. i_{bdi} and i_{bqi} are the direct and quadrature output current components connected to the bus. v_{bdi} and v_{bqi} are the direct and quadrature components of output voltages.

The action of the agent is the adjustable value of the secondary voltage, which ranges from 1.00 pu to 1.14 pu. The action space is designed as a discrete action, and the joint action is $A_t = v_{n1} \times v_{n2} \times \dots \times v_{nN}$. The design of the reward function is closely related to voltage stability. The higher the reward, the faster it quickly returns to the reference value.

$$r_{i,t} = \begin{cases} 0.05 - |1 - v_i|, & v_i \in [0.95, 1.05] \\ -|1 - v_i|, & v_i \in [0.8, 0.95] \cup v_i \in [1.05, 1.25] \\ -10, & \text{violate} \end{cases} \quad (4)$$

3.1. Multi-Agent Advantage Actor-Critic

Consider a distributed energy system of N agents operating in a joint environment. Since the energy nodes are not entirely observable, the environment is constructed as a partially observable Markov decision process [29]. It is formally defined as a tuple: $\langle N, S, A, P, R, O, \gamma \rangle$. The model of the system is denoted as a multi-agent network: $G = (V, \epsilon)$,

where N is the number of agents, and $\varepsilon_{ij} \in \varepsilon : (i, j \in N, i \neq j)$ is the edge of agent i and agent j . Each agent can share information with its neighbors. S, O and $A = [A_1, A_2 \dots A_N]$ are the state space, observation space and action space, respectively. In this paper, the action space is the secondary voltage set point of each energy node, which is discrete. The state transition function is expressed as: $P(s' | s, a) : S \times A \times S \rightarrow [0, 1]$. In our design, each agent shares the global reward $R = O \times A$. γ is the discount factor.

An independent Q-learning algorithm (IQL) is a typical multi-agent reinforcement learning algorithm. Other agents are directly regarded as part of the environment; that is, each agent is solving a single agent task [30]. The environment is not stable because of the existence of other agents. This will lead to the calculations not converging, and the agent will easily fall into endless exploration. Considering these problems, the IQL algorithm was combined with the actor-critic method to adaptive traffic signal control, which is called the multi-agent advantage actor-critic(MA2C) algorithm [29,30]. The framework of the training method is shown in Figure 3.

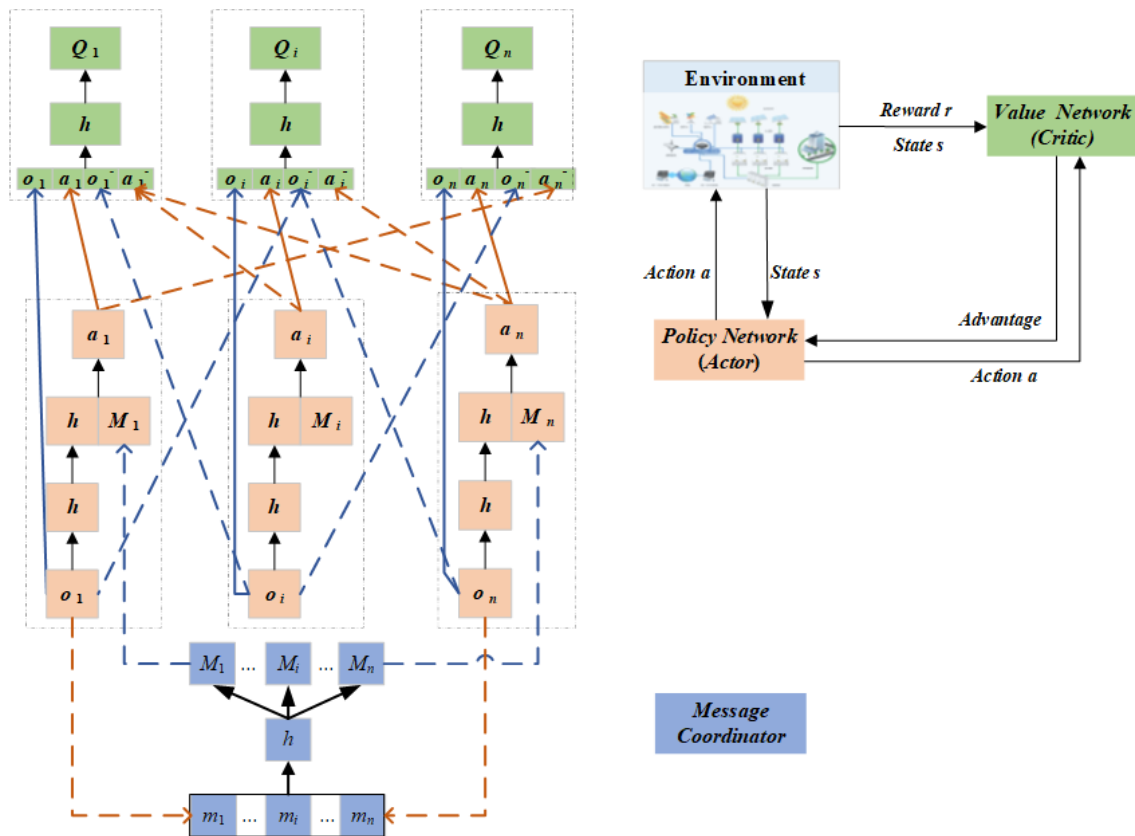


Figure 3. The structure of the multi-agent A2C algorithm with attentional mechanism.

Assume that $\pi_{i,t} = \pi_{\theta_i^-}(\cdot | s_t, \pi_{t-1,j})$, where $\pi_{i,t}$ is the local policy of agent i , and $\pi_{t-1,j}$ is the latest strategy of the neighbor agent. The global reward of agent i is defined as

$$\tilde{R}_{i,t} = \hat{R}_{i,t} + \gamma^{t-T} V_{\omega_i}^-(\tilde{s}_{t_B}, V_i, \pi_{t_T-1, j} | \pi_{\theta_i^-}) \quad (5)$$

where $\hat{R}_{i,t} = \gamma r_i$ is the estimated reward of each energy node. $V_{\omega_i}^-$ is the value regressor. Then, the local advantage function is defined as

$$A_{\theta}^i(s, a) = Q_{\pi}(s, a) - V_{\omega_i}(s, a^-) \quad (6)$$

$$V_{\omega_i}(s, a^-) = \sum_{a^i \in A^i} \pi_{\theta_i}^i(s, a^i) \cdot Q_{\theta}(s, a^i, a^{-i}) \quad (7)$$

The actor-critic algorithm contains two steps in the training process: strategy evaluation and strategy promotion. The action-value TD-learning method is used to update the parameter of policy evaluation in the critic step. Each energy node updates its policy via the policy gradient algorithm in the actor step [21]. The loss function of value and policy for updating is illustrated as

$$L(\omega_i) = -\frac{1}{2T} \sum_{t \in T} (R_{i,t} - V_{\omega_i}(\tilde{s}_t, V_i, \pi_{t-1,j}))^2 \quad (8)$$

$$L(\theta_i) = -\frac{1}{T} \sum_{t \in T} (\log \pi_{\theta_i}(u_{i,t} | \tilde{s}_t, V_i, \pi_{t-1}, N_i) \tilde{A}_{i,t} - \beta \sum_{u_i \in U_i} \log \pi_{\theta_i}(u_{i,t} | \tilde{s}_t, V_i, \pi_{t-1,j})) \quad (9)$$

3.2. MA2C-Attention Algorithm

The secondary voltage control of energy nodes based on multi-agent reinforcement learning requires good coordination and communication among energy nodes. There are usually two training frameworks for solving multi-agent problems based on reinforcement learning, namely centralized and decentralized training frameworks. The centralized training and decentralized execution scheme has become the mainstream alternative in recent years. During training, the centralized mode is adopted. Each agent only makes decisions by applying the trained strategy network according to its own local observations.

Attention mechanism was initially proposed in the field of visual images and has been widely used in many artificial intelligence algorithms [38]. In this paper, soft attention is extended to the multi-agent advantage actor-critic algorithm. The normalized importance score W^k is defined as

$$W^k = \frac{\exp(f(T, S_k))}{\sum_{i=1}^K \exp(f(T, S_i))}, \sum_{i=1}^K W^k \equiv 1 \quad (10)$$

where $f(T, S_k)$ is the defined function to judge the importance score. The input vector $[S_1, S_2, \dots, S_k, \dots, S_K]$ contains the important information, and T is the target vector.

The motivation for introducing attention mechanisms is to obtain a better reward by selectively paying attention to the actions of other energy node. The structure of a multi-agent A2C algorithm with an attentional mechanism is shown in Figure 3. As shown in the figure, each agent contains three types of networks: an actor network, a critic network and a communication network. h is the hidden layer of the deep neural network, m_i is the local message and M_i is the global message. Due to the fully connected network treating all the messages equally, the attention is embedded in the actor network to select more important messages from $m_1, m_2, \dots, m_N$ adaptively. The updating method of the message coordinator network is described as follows. Firstly, we need to set a “query” feature space $m_i^q = m_i \times w_i^q$ and a “key” feature space $m_{ij}^k = m_j \times w_i^k$. The importance score is defined as

$$W_i^j = (m_i^q)^T (m_{ij}^k), i \neq j \quad (11)$$

$$W_i = [W_i^1, \dots, W_i^j, \dots, W_i^N] = \text{softmax}(W_i^1, \dots, W_i^j, \dots, W_i^N) \quad (12)$$

$$\sum_{j=1}^N W_i^j = 1, i \neq j \quad (13)$$

Then the global message is generated based on the importance score

$$M_i = \sum_{j=1}^N W_i^j m_j \quad (14)$$

where W_i^j is the weight of m_j for m_i . The more important local messages will be selected, and the unimportant messages will be ignored according to the weights.

The attention mechanism is combined with the original network; therefore, the training method inherits the above method according to Equations (8) and (9). The detailed illustration of the proposed algorithm is shown in Algorithm 1. The input hyperparameters include $\alpha, \gamma, \eta_\omega, \eta_\theta$, and L, S , where α is the temperature parameter of entropy, γ is the time-discount factor, η_ω is the learning rate of the actor network, η_θ is the learning rate of the critic network, L is the time step of every episode, and S is the total training episodes. The training process of the algorithm circulates thousands of episodes (Lines 2–21). In every episode, each agent picks up information according to the observation information and sends their message to the message coordinator network (Lines 4–6). By setting a “query” feature space and a “key” feature space, we can calculate the importance score. All agents interact with the environment by selecting and executing actions based on the current state and global message (Lines 7–10). Then, the agents move to the next state and receive an immediate reward. The weights of the actor network and the critic network are updated using a policy gradient algorithm (Lines 18–19).

Algorithm 1 MA2C-Attention algorithm

Input: $\alpha, \gamma, \eta_\omega, \eta_\theta, L, S$

Output: θ_i, ω_i

```

1: initialize  $s_0, \pi_{-1}, t \leftarrow 0, D = \emptyset$ ;
2: for  $j = 0$  to  $S - 1$  do
3:   for  $t = 0$  to  $L - 1$  do
4:     for  $i \in V$  do
5:       send  $m_{i,t} = o_{i,t}$ 
6:       send  $M_{i,t} = m_{i,t}$ 
8:     end
9:     for  $i \in V$  do
10:      let  $m_i^q = m_i \times w_i^q, m_{ij}^k = m_j \times w_i^k$ ,
11:      calculating the importance score: Equation (11)
12:      send  $M_{i,t}$  to  $m_{i,t}$ 
13:      update  $a_{i,t} = \pi_{i,t}$ 
14:    end
15:    update  $v_{i,t} = V_{\omega_i}(o_{i,t}, a_{N_i,t})$ 
16:    simulate  $s_{t+1}, r_{i,t}$ 
17:    update  $t \leftarrow t + 1, j \leftarrow j + 1$ 
18:    if  $DONE$  then
19:      for  $i \in V$  do
20:        update  $\theta_i \leftarrow \theta_i + \eta_{\theta_i} \eta_{\theta_i} \nabla_{\theta_i} J(\theta_i)$ 
21:        update  $\omega_i \leftarrow \omega_i + \eta_{\omega_i} \eta_{\omega_i} \nabla_{\omega_i} J(\omega_i)$ 
22:      end
23:    end
24:    initialize  $D \leftarrow \emptyset$ 
25:  end
26:  update  $s_0, \pi_{-1}, t \leftarrow 0$ 
27: end
  
```

4. Simulation

In this section, an easy distributed energy system is built around a distribution network consisting of six buses connected to three renewable energy generators, one distributed energy storage unit, one thermal generator, and one balancing generator. The topology is shown in Figure 4. Each energy was regarded as an agent, and each node can communicate with its neighbor nodes. The secondary control corrects voltage deviations caused by droop control based on the proposed method. The simulation was performed on a Windows 10 server with an Intel(R) Core i5-10400F CPU @ 2.90 GHz processor and 64 GB memory.

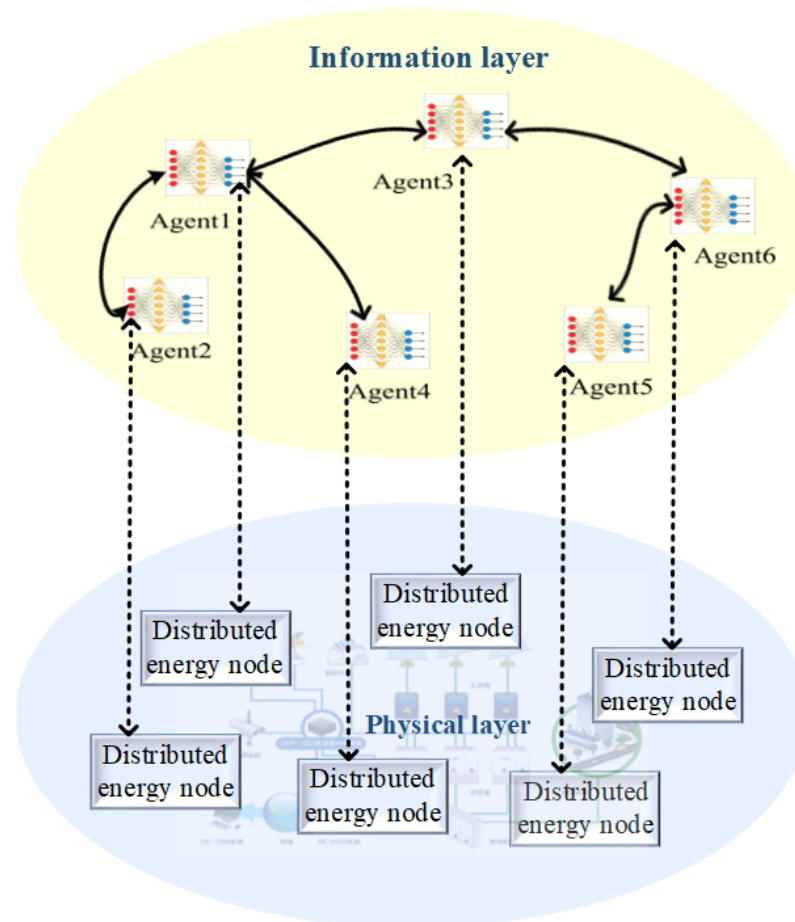


Figure 4. The topological structure of distributed energy system.

In order to illustrate the effectiveness of the proposed algorithm, we compared the algorithm with the independent advantage actor-critic algorithm and multi-agent advantage actor-critic algorithm. We set 4000 training episodes, and one episode lasts 20 steps. During the training process, each episode generated different random seeds. To ensure the fairness of comparison, the different algorithms shared the same random seed. The average reward value of different algorithms is shown in Figure 5.

The red curve is the average reward value trained based on the proposed method. The blue and green curve is the average reward value trained based on MA2C algorithm and IA2C algorithm, respectively. Obviously, the multi-agent advantage actor-critic algorithm with attention mechanism has a higher episode reward and faster convergence speed compared to other algorithms.

In order to further verify the effectiveness of the proposed method, we conducted random tests on the trained model. The test results in different random seeds are shown in Table 1. The average reward of the MA2C-attention algorithm is 0.264. The average reward of five groups was tested, and the results obtained are 1.056 and 1.58 times compared with the MA2C and IA2C algorithm.

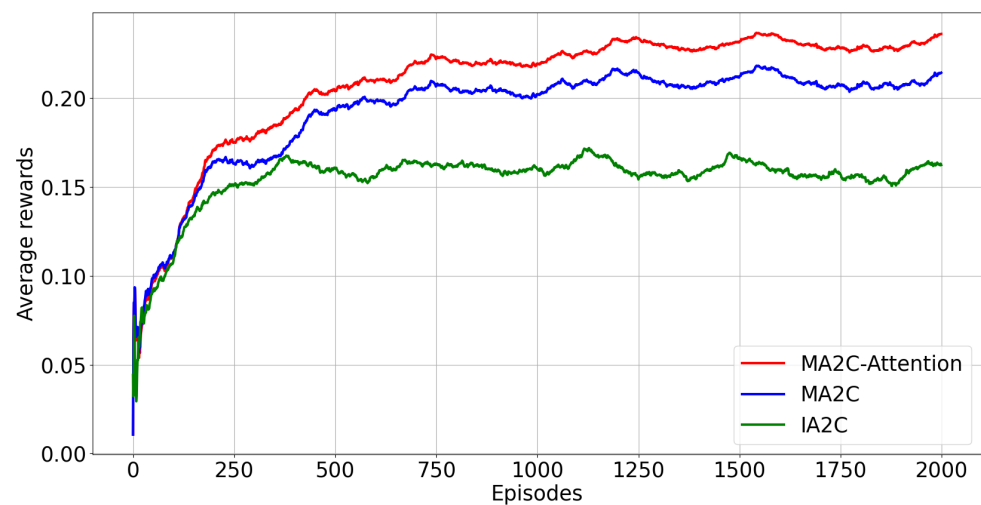


Figure 5. The average reward value under different algorithms.

Table 1. Average reward value under random test.

Random Seeds	MA2C-Attention	MA2C	IA2C
Case 1	0.27	0.26	0.178
Case 2	0.25	0.25	0.112
Case 3	0.26	0.25	0.194
Case 4	0.27	0.26	0.177
Case 5	0.27	0.25	0.175
Average reward	0.264	0.25	0.167

We also compared the voltage control results between the secondary controller and only the primary controller. The voltages of the six distributed generators under different algorithms are shown in Figure 6. The abscissa value 0–5 represents the terminal voltages of the six distributed generators. The ordinate represents the unit value of the voltage. The specific voltage value of Figure 6 is shown in Table 2. Agent 1–6 denotes the six energy nodes, respectively. According to the chart, we can clearly see that the voltage value of each node is closer to the reference value under the proposed method. The advantages of the method proposed in this paper are more clear winners than the typical droop control. Due to the randomness of renewable energy, under primary control, the voltage cannot recover to the reference value rapidly. The proposed algorithm can quickly promote the node voltage to the reference value. The voltage fluctuations within the allowable range and remain relatively stable.

Table 2. Voltage value under different algorithms.

Method	MA2C-Attention	MA2C	IA2C	Droop Control
Agent 1	1.00975192	1.00467732	0.97759587	0.9
Agent 2	1.00788861	0.9908174	0.98287053	0.89002
Agent 3	0.99718829	0.9887299	0.97389743	0.90125
Agent 4	0.99951272	0.9894690	0.97161282	0.89234
Agent 5	0.99088914	0.9894690	0.98344116	0.91232
Agent 6	0.99591974	0.9862523	0.9800902	0.90023

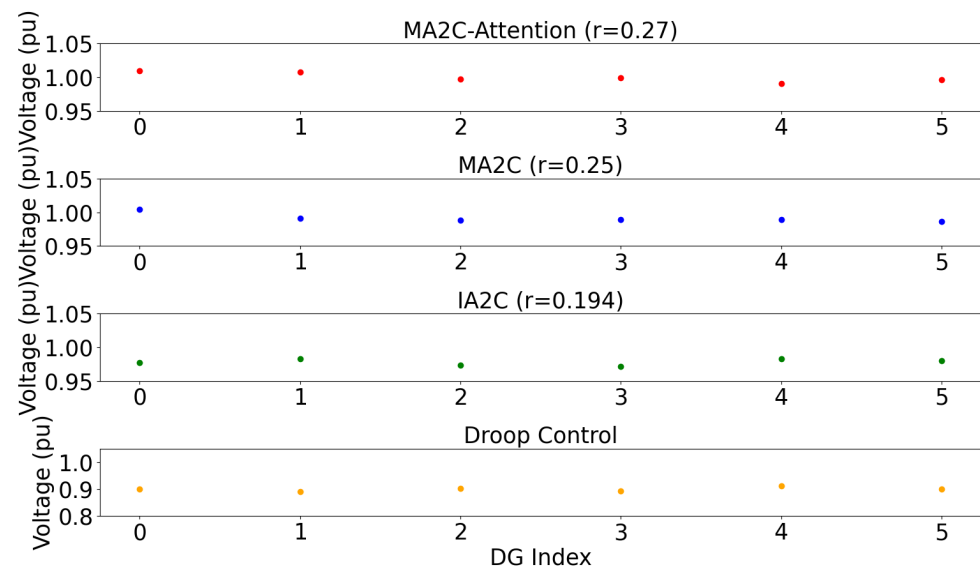


Figure 6. The voltages distributions under different algorithms.

5. Conclusions

This paper proposes a multi-agent advantage actor-critic algorithm to solve the collaborative control of secondary voltage in a distributed energy system. The attentional mechanism is extended to the multi-agent advantage actor-critic algorithm to improve communication efficiency, so that all agents cooperate more effectively with each other. The simulation results of a simple distributed energy system demonstrate that the proposed algorithm achieves the cooperative control of secondary voltage. Compared with the independent advantage actor-critic algorithm and the multi-agent advantage actor-critic algorithm, the introduction of the attentional mechanism facilitates the promotion of global reward. This study has potential limitations. The construction of the distributed energy system lacks the consideration of an actual power system. We took into account the effect of voltage offset in the design of the algorithm's model, but the frequency offset is also an important influencing factor for the stability of the distributed energy system. We will build a realistic simulation environment through the collection and analysis of the actual operating system. In addition, we will consider the design of the three-level control scheme of the distributed energy system. The objective function will be designed according to the needs of the user to achieve optimal control of the distributed energy system. The dispatch scheme will be designed based on the multi-agent reinforcement learning algorithm to provide the reference value of the optimal power generation of each distributed energy. The reference values transmitted to the lower layer through the communication network control the output power of energy nodes to realize the safe and economic operation of the distributed energy system.

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

Funding: This research was funded by Science and Technology Project of State Grid Corporation of China (KJ22-1-63: Research on collaborative optimization method of multi-agent reinforcement learning in distributed information energy system).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wang, X. Discussion on Problems Related to Development of the Distributed Energy System. *Electr. Power Technol. Econ.* **2007**, *19*, 6–8.
- González-Torres, M.; Pérez-Lombard, L.; Coronel, J.F.; Maestre, I.R.; Yan, D. A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Rep.* **2022**, *8*, 626–637. [\[CrossRef\]](#)
- Shi, Z.; Wang, W.; Huang, Y.; Li, P.; Dong, L. Simultaneous optimization of renewable energy and energy storage capacity with the hierarchical control. *CSEE J. Power Energy Syst.* **2022**, *8*, 95–104.
- Huang, X.; Liu, Y.; Liao, Y.; Jiang, Z.; He, J.; Li, Y. Modeling of distributed energy system with multiple energy complementation. In Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 20–22 October 2018; pp. 1–6.
- Li, X.; Ma, R.; Yan, N.; Wang, S.; Hui, D. Research on Optimal Scheduling Method of Hybrid Energy Storage System Considering Health State of Echelon-Use Lithium-Ion Battery. *IEEE Trans. Appl. Supercond.* **2021**, *31*, 1–4. [\[CrossRef\]](#)
- Li, X.; Wang, L.; Yan, N.; Ma, R. Cooperative Dispatch of Distributed Energy Storage in Distribution Network With PV Generation Systems. *IEEE Trans. Appl. Supercond.* **2021**, *31*, 1–4. [\[CrossRef\]](#)
- Liu, X.; Liu, Y.; Liu, J.; Xiang, Y.; Yuan, X. Optimal planning of AC-DC hybrid transmission and distributed energy resource system: Review and prospects. *CSEE J. Power Energy Syst.* **2019**, *5*, 409–422. [\[CrossRef\]](#)
- Liu, Y.; Li, Y.; Gooi, H.B.; Jian, Y.; Xin, H.; Jiang, X.; Pan, J. Distributed Robust Energy Management of a Multimicrogrid System in the Real-Time Energy Market. *IEEE Trans. Sustain. Energy* **2019**, *10*, 396–406. [\[CrossRef\]](#)
- Wang, H.; Huang, J. Incentivizing Energy Trading for Interconnected Microgrids. *IEEE Trans. Smart Grid* **2018**, *9*, 2647–2657. [\[CrossRef\]](#)
- Han, Y.; Zhang, K.; Li, H.; Coelho, E.A.A.; Guerrero, J.M. MAS-Based Distributed Coordinated Control and Optimization in Microgrid and Microgrid Clusters: A Comprehensive Overview. *IEEE Trans. Power Electron.* **2018**, *33*, 6488–6508. [\[CrossRef\]](#)
- Kovaltchouk, T.; Blavette, A.; Aubry, J.; Ahmed, H.B.; Multon, B. Comparison Between Centralized and Decentralized Storage Energy Management for Direct Wave Energy Converter Farm. *IEEE Trans. Energy Convers.* **2016**, *31*, 1051–1058. [\[CrossRef\]](#)
- Mu, C.; Tang, Y.; He, H. Improved Sliding Mode Design for Load Frequency Control of Power System Integrated an Adaptive Learning Strategy. *IEEE Trans. Ind. Electron.* **2017**, *64*, 6742–6751. [\[CrossRef\]](#)
- Mu, C.; Liu, W.; Xu, W. Hierarchically Adaptive Frequency Control for an EV-Integrated Smart Grid With Renewable Energy. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4254–4263. [\[CrossRef\]](#)
- Zhang, Y.; Zhang, J.; Gao, W.; Zheng, X.; Yang, L.; Hao, J.; Dai, X. Distributed electrical energy systems: Needs, concepts, approaches and vision. *Acta Autom. Sin.* **2017**, *43*, 1544–1554.
- Yu, X.; Jiang, Z.; Zhang, Y. Control of Parallel Inverter-Interfaced Distributed Energy Resources. In Proceedings of the 2008 IEEE Energy 2030 Conference, Atlanta, GA, USA, 17–18 November 2008; pp. 1–8.
- Gholami, S.; Aldeen, M.; Saha, S. Control Strategy for Dispatchable Distributed Energy Resources in Islanded Microgrids. *IEEE Trans. Power Syst.* **2018**, *33*, 141–152. [\[CrossRef\]](#)
- Meng, X.; Liu, J.; Liu, Z. A Generalized Droop Control for Grid-Supporting Inverter Based on Comparison Between Traditional Droop Control and Virtual Synchronous Generator Control. *IEEE Trans. Power Electron.* **2019**, *34*, 5416–5438. [\[CrossRef\]](#)
- Foo, Eddy, Y.S.; Gooi, H.B.; Chen, S.X. Multi-Agent System for Distributed Management of Microgrids. *IEEE Trans. Power Syst.* **2015**, *30*, 24–34. [\[CrossRef\]](#)
- Zhao, T.; Ding, Z. Distributed Finite-Time Optimal Resource Management for Microgrids Based on Multi-Agent Framework. *IEEE Trans. Ind. Electron.* **2018**, *65*, 6571–6580. [\[CrossRef\]](#)
- Singh, V.P.; Kishor, N.; Samuel, P. Distributed Multi-Agent System-Based Load Frequency Control for Multi-Area Power System in Smart Grid. *IEEE Trans. Ind. Electron.* **2017**, *64*, 5151–5160. [\[CrossRef\]](#)
- Colson, C.M.; Nehrir, M.H. Comprehensive Real-Time Microgrid Power Management and Control With Distributed Agents. *IEEE Trans. Smart Grid* **2013**, *4*, 617–627. [\[CrossRef\]](#)
- Mu, C.; Zhang, Y.; Jia, H.; He, H. Energy-Storage-Based Intelligent Frequency Control of Microgrid With Stochastic Model Uncertainties. *IEEE Trans. Smart Grid* **2020**, *11*, 1748–1758. [\[CrossRef\]](#)
- Ju, L.; Zhang, Q.; Tan, Z.; Wang, W.; He, X.; Zhang, Z. Multi-agent-system-based coupling control optimization model formicro-grid group intelligent scheduling considering autonomy-cooperative operation strategy. *Energy* **2018**, *157*, 1035–1052. [\[CrossRef\]](#)
- Morstyn, T.; Hredzak, B.; Agelidis, V.G. Cooperative Multi-Agent Control of Heterogeneous Storage Devices Distributed in a DC Microgrid. *IEEE Trans. Power Syst.* **2016**, *31*, 2974–2986. [\[CrossRef\]](#)
- Gao, Y.; Wang, W.; Yu, N. Consensus Multi-Agent Reinforcement Learning for Volt-VAR Control in Power Distribution Networks. *IEEE Trans. Smart Grid* **2021**, *12*, 3594–3604. [\[CrossRef\]](#)
- Liu, W.; Gu, W.; Sheng, W.; Meng, X.; Wu, Z.; Chen, W. Decentralized Multi-Agent System-Based Cooperative Frequency Control for Autonomous Microgrids With Communication Constraints. *IEEE Trans. Sustain. Energy* **2014**, *5*, 446–456. [\[CrossRef\]](#)
- Mu, C.; Peng, J.; Tang, Y. Learning-based control for discrete-time constrained nonzero-sum games. *CAAI Trans. Intel. Technol.* **2020**, *6*, 203–213. [\[CrossRef\]](#)
- Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [\[CrossRef\]](#)

29. Zhang, K.; Yang, Z.; Liu, H.; Zhang, T.; Basar, T. Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 5872–5881.
30. Sun, C.; Mu, C. Important Scientific Problems of Multi-Agent Deep Reinforcement Learning. *Acta Autom. Sin.* **2020**, *46*, 1301–1312.
31. Wang, J.; Xu, W.; Gu, Y.; Song, W.; Green, T.C. Multi-agent reinforcement learning for active voltage control on power distribution networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3271–3284.
32. Mao, H.; Zhang, Z.; Xiao, Z.; Gong, Z.; Ni, Y. Learning agent communication under limited bandwidth by message pruning. In Proceedings of the AAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 5142–5149.
33. Sukhbaatar, S.; Fergus, R.; Learning multiagent communication with backpropagation. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.
34. Chen, D.; Chen, K.; Li, Z.; Chu, T.; Yao, R.; Qiu, F.; Lin, K. PowerNet: Multi-Agent Deep Reinforcement Learning for Scalable Powergrid Control. *IEEE Trans. Power Syst.* **2022**, *37*, 1007–1017. [[CrossRef](#)]
35. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1086–1095. [[CrossRef](#)]
36. Lowe, R.; Wu, Y.I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
37. Tampuu, A.; Matiisen, T.; Kodelja, D.; Kuzovkin, I.; Korjus, K.; Aru, J.; Aru, J.; Vicente, R. Multiagent cooperation and competition with deep reinforcement learning. *PLoS ONE* **2017**, *12*, e0172395. [[CrossRef](#)] [[PubMed](#)]
38. Mao, H.; Zhang, Z.; Xiao, Z.; Gong, Z.; Ni, Y. Learning multi-agent communication with double attentional deep reinforcement learning. *Auton. Agents Multi-Agent Syst.* **2020**, *34*, 1–34. [[CrossRef](#)]