frontiers | Frontiers in Energy Research

ORIGINAL RESEARCH
published: 04 July 2022
doi: 10.3389/fenrg.2022.933011

Check for updates

Security Constrained Dispatch for Renewable Proliferated Distribution Network Based on Safe Reinforcement Learning

Han Cui[1,2], *Yujian Ye*[2,3]*, *Qidong Tian*[4] *and Yi Tang*[1,3]

[1]*School of Cyber Science and Engineering, Southeast University, Nanjing, China,* [2]*Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing, China,* [3]*School of Electrical Engineering, Southeast University, Nanjing, China,* [4]*Shenzhen Power Supply Bureau, China Southern Power Grid, Shenzhen, China*

As the terminal of electricity consumption, the distribution network is a vital field to lower the carbon emission of the power system. With the integration of distributed energy resources, the flexibility of the distribution network has been promoted significantly where dispatch actions can be employed to lower carbon emissions without compromising the accessibility of reliable electricity. This study proposes a security constrained dispatch policy based on safe reinforcement learning for the distribution network. The researched problem is set up as a constrained Markov decision process, where continuous-discrete mixed action space and high-dimensional state space are in place. In addition, security-related rules are embedded into the problem formulation. To guarantee the generalization of the reinforcement learning agent, various scenarios are generated in the offline training stage, including randomness of renewables, scheduled maintenance, and different load profiles. A case study is performed on a modified version of the IEEE 33-bus system, and the numerical results verify the effectiveness of the proposed method in decarbonization.

Keywords: decarbonization dispatch, active distribution networks, safe reinforcement learning, renewable generation, electricity storage

Edited by:
Hao Wang,
Monash University, Australia

Reviewed by:
Yingjun Wu,
Hohai University, China
Tianxiang Cui,
The University of Nottingham Ningbo (China), China

***Correspondence:**
Yujian Ye
yeyujian@seu.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *30 April 2022*
Accepted: *23 May 2022*
Published: *04 July 2022*

Citation:
Cui H, Ye Y, Tian Q and Tang Y (2022) Security Constrained Dispatch for Renewable Proliferated Distribution Network Based on Safe Reinforcement Learning. Front. Energy Res. 10:933011. doi: 10.3389/fenrg.2022.933011

INTRODUCTION

Decarbonization has been a global consensus to tackle climate change (Ou et al., 2021), which has promoted the prosperity of renewable energy sources (RES) in the past years. Moreover, RES will take the dominant share in global electricity generation, increasing from 29% in 2020 to over 60% in 2030 and to nearly 90% in 2050 (IEA, 2021). As in a distribution network, a massive influx of distributed RES would help approach *NET ZERO* (Ahmed et al., 2022). However, the output of renewable energy has the inherent characteristics of uncertainty and variability, introducing lots of difficulties to distribution network dispatches, such as real-time power generation and consumption imbalance, voltage fluctuation, frequency oscillation, and transmission congestion (Bistline, 2021; Abd El-Kareem et al., 2021; Husin and Zaki, 2021). Therefore, a dispatch strategy for a distribution network with a high percentage of RES is required to handle uncertainty and variability caused by RES.

To address this issue, various methods have been proposed from two aspects: model-based optimization and data-driven method, e.g., reinforcement learning and deep learning. For the former category, the optimization problem can be formulated from the aspect of either electricity market, where price signals would affect the decision-making process of participants (Allan et al., 2015; Lin et al., 2017; Ye et al., 2019), or centralized operation, where dispatch center would make all dispatch actions according to its objectives and real-time observations (Wang et al., 2018). In Caramanis et al. (2016), a centralized optimization problem is formulated to discover the electricity pricing strategy so that dispatchable resources can be scheduled efficiently. Peer-to-peer is another form of market that governs the distributed network, and the alternating direction method of multipliers (ADMM) is applied Nguyen (2020) to find the optimal energy management strategy by peer trading. To further promote the scale of the researched problem, distributed ADMM is applied to handle the multiple microgrids situation where DERs. A bi-level trading strategy is developed to coordinate the microgrids and distribution network so that power supply and consumption can be balanced economically (Wang et al., 2019). A study conducted by Hu et al. (2018) have the similar carbon emission reduction target to our research; however, the dispatch is focused on the interaction between the transmission network and distribution network, where transmission power, locational marginal emission (LME), and locational marginal price (LMP) are iterated to decrease the emission and transmission cost.

However, this type of approach exhibits a significant drawback: frequent changes in the distribution network cause the employed system model for optimization become inaccurate, which deteriorates the effectiveness of the dispatch decisions. Furthermore, methods belonging to stochastic programming (SP) would lead to a significant computational burden due to the increasing scale of the power grid. Alternatively, robust optimization (RO) approaches may be over-conservative driven by their nature in hedging against the worst-case realization of the uncertainties. In Zhou et al. (2019), a decentralized dispatch framework is proposed to handle the power fluctuation caused by renewables, where robustness is realized by a column-and-constraint generation algorithm. In a study conducted by Zhang et al. (2018), to overcome the conservativeness of robust optimization, extreme cases are fetched from historical data instead of generating a large simulated dataset. The selection method of extreme cases is theoretically proven to be robust under all potential situations.

Finally, large-scale model-based optimization is characterized by significant non-linearity which leads to solution inaccuracy, and the long period of calculation leaves the time window to execute the control action too short to catch up with real-time situations.

With the development of artificial intelligence, reinforcement learning (RL) algorithms have shown great advantages in real-time policy. RL agent optimizes its policy in the extensive interaction between environments, where policy is updated to maximize the reward. By setting up a comprehensive dataset in the environment, the RL agent would learn to handle all possible scenarios, which ensures policy adjusts to the uncertainties of RES and the real-time status of the distribution network (Al-Saffar and Musilek, 2021; Cao et al., 2021; Li et al., 2021; Zhang et al., 2021). In Cao et al. (2021), proximal policy optimization (PPO) is applied to a distribution network to absorb the power flow fluctuation caused by renewables, where storage devices can be controlled discretely. Alternatively, a deep deterministic policy gradient (DDPG) deals with continuous action space. In the work of Zhang et al. (2021), voltage drift problems caused by the randomness of renewable are solved by DDPG, where static var compensators are controlled continuously to keep the voltage at each bus within the permitted range.

Despite the significant application potential, the examined problem features a mixed discrete (e.g., topology switching) and continuous (e.g., electricity storage) action space, whereas previous RL methods can only handle either discrete or continuous action spaces. Furthermore, the examined problem dictates that the dispatch actions need to respect the distribution network constraints, e.g., actions on the EV charging could lead to low voltage at the access point. Thus, constraint satisfaction must be accounted for during policy learning. Based on these considerations, interior-point policy optimization (IPO) (Liu et al., 2020), a safe RL algorithm, is applied to the distribution network.

To absorb the uncertainty and variability, various types of dispatchable resources are utilized to optimize the operation status of the distribution network, including distributed generator, grid topology, responsive load, and electricity storage (Bizuayehu et al., 2016; Ju et al., 2016; Ghasemi and Enayatzare, 2018; Arfeen et al., 2019; Mohammadjafari et al., 2020). With an appropriate control strategy, local residual power can be consumed, stored, or transmitted, while power shortage can be compensated by electricity storage, distributed generator, flexible load, or transmission network. To achieve the long-term target of *NET ZERO*, a study conducted by Pehl et al. (2017) has analyzed the life-cycle carbon emission of power system components. From the economic cost perspective, a study conducted by Brouwer et al. (2016) proposed several scenarios for reducing carbon emissions by up to 96% with the integration of intermittent renewables. Consequently, a combination of various dispatchable resources enables the distribution network to operate in more reliable and environmental-friendly manner.

The contribution of this study is listed as follows:

1) The decarbonization dispatch problem is formulated as a Constrained Markov Decision Process (CMDP), which provides the foundation for the RL method
2) Minimize the carbon emission in a distribution network without violating power system security rules, providing guidance for future power system operation
3) The proposed algorithm dispatches different types of resources and the continuous-discrete mixed actions can handle different scenarios smoothly

The rest of the study is organized as follows. *Introduction* formulates the distribution network dispatch problem in detail and introduces related features of various dispatchable resources. It presents the dispatch method based on safe RL and clarifies the mechanism of related algorithms. It demonstrates the numerical test results of the proposed method on the IEEE 33-bus system. Eventually, the presented work is summarized in *Introduction*.

PROBLEM FORMULATION

With the proliferated renewables becoming important power sources in the distribution network, it would be necessary to perform dispatch actions on DERs to overcome the variable power supply from renewables (Huang et al., 2019). Different from existing research that quantifies the dispatch effect according to economic cost, carbon emission is emphasized in this study with the premise of reliable electricity supply. By consuming the electricity generated by a wind farm or solar plant, carbon emission is minimized. Dispatch resources contain network topology, controllable load, distributed generators, and electricity storage. The optimization target is to find a policy that minimizes the carbon emission at the prerequisites of meeting all constraints.

In this study, the decarbonization-driven dispatch of the distribution network is formulated as CMDP, which can be expressed with a tuple (S, A, P, R,γ, C), where S is the set of state variables in the distribution network; A represents the set of dispatch actions; P is the transition probability function between states; R is the reward function during state transition; γ is a discounted factor for the reward at different time steps; C is constraints that related to the security of distribution network.

State Space

In this problem, the measurement of components in the distribution network constitutes the state space: power from the external power grid P^{ex}, power of generators, wind/solar farm P^g, power of electricity storage P^s, State of Charge (SOC) B^s, load consumption P^l, switch status W. At each time step, these variables reveal the real-time situation of the distribution network, which is the foundation for dispatch. Let S_t be the state vector at time step t.

$$\boldsymbol{S_t} = \left(\boldsymbol{P^{ex}_{1:\,N_E}}, \boldsymbol{P^{g}_{1:\,N_G}}, \boldsymbol{P^{s}_{1:\,N_S}}, \boldsymbol{B^{s}_{1:\,N_S}}, \boldsymbol{P^{l}_{1:\,N_L}}, \boldsymbol{W_{1:N_w}}\right),$$

where N_E, N_G, N_S, N_L, N_W represent the total number of external grid interfaces, generators, storage units, loads, and switches in the distribution network. Among these state variables, $\boldsymbol{W_{1:\,N_w}}$ is the only discrete type with a possible value of 0/1, which corresponds to the disconnection/connection status of switches. In addition, the large influx of renewables has made the $\boldsymbol{P^{g}_{1:\,N_G}}$ different from that of traditional distribution networks in the aspects of randomness and dimensionality.

Action Space

Previous dispatch strategy has predominately focused on the transmission network, where large power plants can be used to improve the power flow distribution. Meanwhile, as the affiliate of the transmission network, the distribution network can also benefit from those dispatch actions. However, DERs have changed the situation where even if the high-voltage-level power grid operates smoothly, the distribution network could suffer from volatility. Consequently, dispatch actions in the distribution network are necessary to handle the chaos caused by renewables.

In this study, four types of actions are employed, namely generator redispatch a^g, storage unit control a^s, load shedding a^l, and switch control a^T, which is the only discrete action. The continuous-discrete mixed action space enables the dispatch effect with multiple granularities. Let A_t be the action vector at time step t.

$$\boldsymbol{A_t} = \left(\boldsymbol{a^{g}_{1:\,N_g}}, \boldsymbol{a^{s}_{1:\,N_s}}, \boldsymbol{a^{l}_{1:\,N_l}}, \boldsymbol{a^{w}_{1:\,N_w}}\right),$$

where N_g, N_s, N_l, N_w represent the number of dispatchable resources: generator, storage unit, load, and switch.

Physics constraints on these actions are defined as follows:

$$\left|a_i^g\right| \le G_i^{ramp},$$
$$S_i^{min} \le a_i^s \le S_i^{max},$$
$$0 \le a_i^g \le DR_i^{max},$$

where G_i^{ramp} stands for the ramp limit for the generator i, S_i^{min}, and S_i^{max} are the discharging and charging limit for the storage unit i, DR_i^{max} are the maximum power of the controllable load.

Environment

The CMDP problem is established using Python, in which the model of the distribution network is built with Pandapower, and dispatch actions are simulated in Grid2Op. In the environment, power flow calculation can be performed at each time step and dispatch actions are reflected in the real-time model. In the training process of IPO, the dispatch agent interacts with the Grid2Op object, realizing the action of space exploration and fetching the results.

Reward

Since the objective is to minimize carbon emission, the reward is set as the negative number of total carbon emissions.

$$R = -\left(\sum_{i=1}^{N_E} \rho_i^{ex} P_i^{ex} + \sum_{i=1}^{N_G} \rho_i^{g} P_i^{g} + \sum_{i=1}^{N_L} \rho_i^{l} P_i^{l}\right),$$

where $\rho_i^{ex}, \rho_i^{g}, \rho_i^{l}$ is the carbon emission coefficient for the external grid, generators, and load. In practice, the electricity storage unit also has carbon emissions due to the loss of charging/discharging, which is neglected in this study due to its relatively minor impact. Among ρ_i^g, carbon emission coefficient varies due to the type of generators. For example, distributed generators that consume fossil fuels produce lots of carbon dioxide, while wind turbines and solar panels have no greenhouse gas emissions. The reward has a theoretical upper bound 0, which means no carbon emission, which cannot be reached in this study due to the dependency on distribution generators and the external grid. Generally, the reward can be maximized by full usage or storage of electricity generated by renewables. The target of the dispatch policy π is to strive for the discounted return as high as possible.

$$J(\pi) = E_{\tau\sim\pi}\left[\sum_{t=0}^{T} \gamma^t R_t\right],$$

where γ is the discount factor that sums up the time series, τ is the state-action trajectory of T time steps.

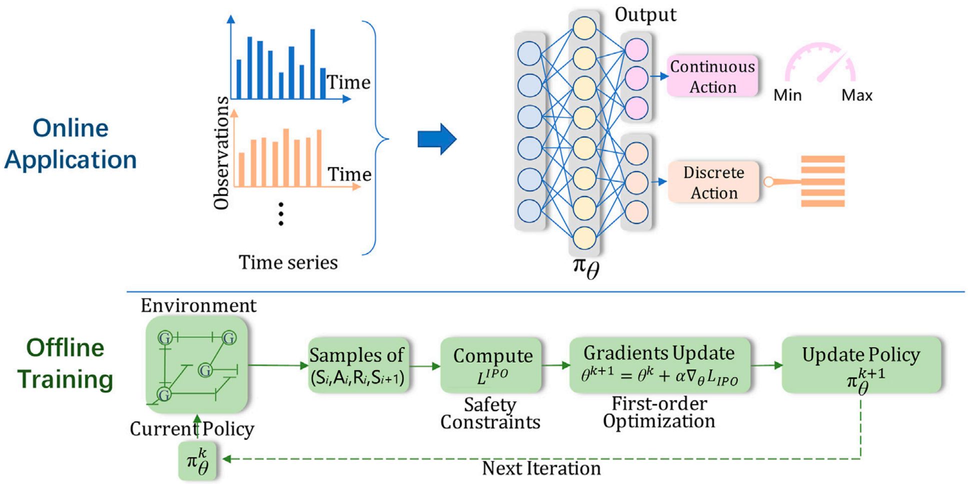

FIGURE 1 | Framework of the proposed method.

Safety Constraints

The power system is essential to modern society so safe and reliable electricity access is critical. It is necessary to consider safety constraints in the dispatch. In this study, constraints related to the storage unit, switches and voltage are considered. For the storage unit, minimum electricity storage is required to provide an emergency reserve. For switches, topology modification must not form an isolated grid. For each bus at the distribution network, the voltage must within a reasonable range. All three constraints are defined as follows:

$$0.2 \le \boldsymbol{B}^s_{1:N_S} \le 0.8,$$
$$G \equiv 1,$$
$$0.95 \le \boldsymbol{V}_{1:N_B} \le 1.05,$$

where G is the number connected graph, V is the voltage at buses, and N_B is the total number of buses. To quantify the violation of safety constraints, the cost function is defined as

$$c_t = \Delta \boldsymbol{B}^s_{1:N_S} + \Delta \boldsymbol{V}_{1:N_B},$$

which represents the SOC and voltage deviation from the safety range. The isolated grid constraint is individually guaranteed by checking the status of switches, which cannot be compromised in any scenario. Consequently, safety constraints can be written in a similar format as a reward.

$$J_C(\pi) = E_{\tau\sim\pi}\left[\sum\nolimits_{t=0}^{T} \gamma^t c_t\right],$$
$$J_C(\pi) \le \varepsilon,$$

where ε is the tolerance for safety constraints.

PROPOSED METHOD

For general reinforcement learning problems, constraints are embedded in the environment that all action exploration is reasonable. For example, in the "inverted pendulum", no matter what action is taken, the system is safe and intact. However, for learning tasks like power system dispatch, inappropriate action might cause severe damage to people or property. Consequently, artificial rules concerning safety are formulated to address this issue. These rules cannot be explicitly executed in action space because whether the rules are breached needs to be judged based on both action and current state. It would be a heavy computational burden to do this judgment for the whole action space before decision. This type of problem is characterized as a safety-related reinforcement learning problem, where safe RL performs well than traditional RL. Algorithms like DDPG cannot solve the CMDP problem directly where safety constraints have to be transformed into a penalty term in reward.

Based on the idea of the interior-point method, the barrier function is used in IPO to quantify the constraint violations. Since the logarithm function has a feature that the value of function approaching negative infinite as variable approaching zero, it is a perfect function to punish the constraint violations. The advantages of IPO are: 1) optimization process of IPO is first-order so that the training efficiency is better than other RL algorithms. 2) multiple safety constraints can be considered in the objective function by simply adding more barrier functions. The IPO can be formulated as

$$\boldsymbol{max}\, \boldsymbol{L}^{CLIP}(\theta) = \boldsymbol{E}_t\left[min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\boldsymbol{A}_t, clip\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}, 1, 1 - \epsilon, 1+\epsilon\right)\boldsymbol{A}_t\right)\right],$$
$$s.t. J_C(\pi) \le \varepsilon,$$

where $\boldsymbol{A}_t$ is the advantage function. Suppose $\widehat{J_C}(\pi) = J_C\pi - \varepsilon$, then $\widehat{J_C}(\pi)$ is a non-positive number if the safety constraints are not violated to the extent of ε. A perfect barrier function is defined as

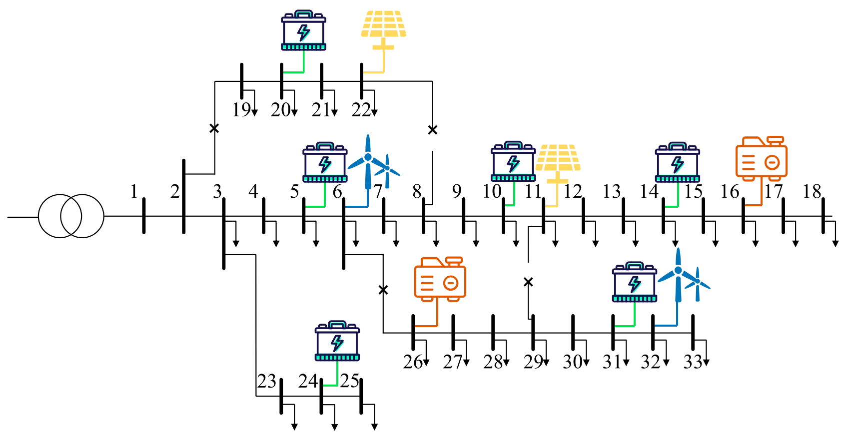

FIGURE 2 | Modified IEEE 33-bus system.

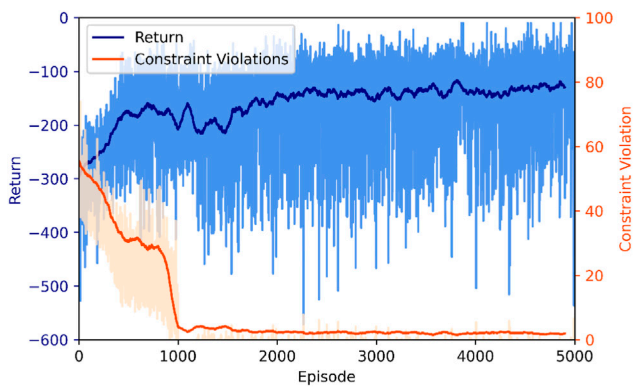

FIGURE 3 | Training process of dispatch agent based on IPO.

$$\mathrm{I}\left(\hat{J}_C(\pi)\right) = \begin{cases} 0, \hat{J}_C(\pi) \le 0, \\ -\infty, \hat{J}_C(\pi) > 0. \end{cases}$$

To find a differentiable function that fits this characteristic, the logarithm function can be applied to $\hat{J}_C(\pi)$.

$$\phi\left(\hat{J}_C(\pi)\right) = \frac{log\left(-\hat{J}_C(\pi)\right)}{k},$$

in which k is a positive hyperparameter. The optimization problem of IPO can be written in a non-constraint format.

$$max\, \boldsymbol{L}^{CLIP}(\boldsymbol{\theta}) + \phi\left(\hat{J}_C(\pi)\right).$$

The whole process of the proposed method can be seen in **Figure 1** where offline training and online application consist of the whole framework. In the training stage, the agent interacts with the distribution model in the Grid2op environment to strengthen its dispatch policy π. Once the training process converges to an acceptable level, the agent can be deployed for an online application. In the online stage, real-time data and historical data are collected and sent to the policy network. After the calculation in the policy neural networks, neurons of the output layer give the suggested action. Ideally, training data enable the agent to be capable of solving various problems, while it might fail in extreme cases. Similar to supervised learning algorithms, it is also necessary to promote the generalization of RL agents.

CASE STUDY

Test Case Preparation

To demonstrate the effectiveness and advantage of the proposed dispatch method, numerical tests are performed on the modified IEEE 33-bus system, as is shown in **Figure 2**. In this test system, dispatchable resources include six electricity storage units, two distribution generators at Bus 26, four switches, and responsive load at Bus 4, 9, 13, 19, 23, and 28. Detailed information on the test system can be seen in **Table 1**. The operation data are simulated for 364-days with 5-min intervals, in which 260 days are used as a training set and 84 days are tested. Each day is seen as an episode of 288 steps for the dispatch agent. All simulations are performed on a server with an NVIDIA 3090Ti GPU and an Intel i7-10700K CPU. The Main Python package used in this research is Pandapower, Grid2op, and Tensorflow.

Evaluation of the Proposed Method

The training process of the IPO agent is shown in **Figure 3** both reward and constraint violation is depicted by the blue curve and red curve respectively. And the moving average of 50 episodes is drawn with a darker color. It can be seen that with the training process continuing, the reward goes up and converge to-245. This trend illustrates the effectiveness of the dispatch policy in

TABLE 1 | Information about the test system.

Distributed generator	Maximum output (MW)	Ramp limit (MW)	ρ (tCO2/MWh)
Bus 26	5	0.2	0.65
Bus 17	3.5	0.15	0.70
Wind farm	Maximum output (MW)	Solar farm	Maximum output (MW)
Bus 6	2	Bus 11	1.5
Bus 32	1.5	Bus 22	1
Storage unit	Maximum charge power (kW)	Maximum discharge power (kW)	SOC(kWh)
5	70	70	700
10	80	80	800
15	80	80	800
20	100	100	1,000
24	80	80	800
31	70	70	700

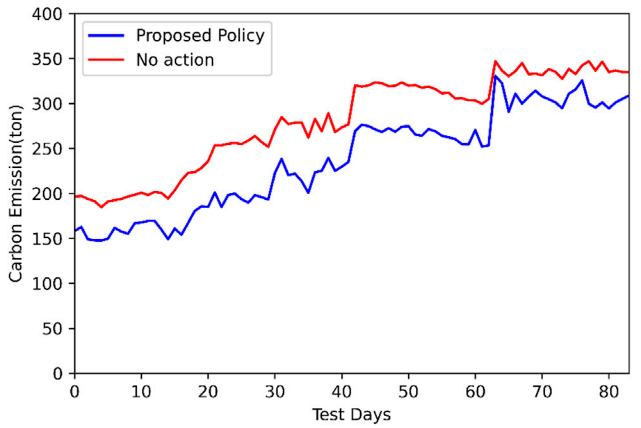

FIGURE 4 | Comparison of the proposed method and no action policy.

reducing carbon emissions. In addition, constraint violation drops dramatically to a small value and converges to the tolerance level, which reveals the advantage of IPO handling the safety constraints.

To show the low-carbon feature of the dispatch policy, a test on the distribution network over an 84-days dataset is performed. Results are shown in **Figure 4**, where the red line represents carbon emission without dispatch actions and the blue line represents the proposed policy. In the no-action cases, the power balance is satisfied by setting the external grid as a slack bus. The carbon emission ranges from 184 to 347 tons without dispatch actions and 147 to 330 with the proposed dispatch policy. The total carbon emission over the 84-day period is 23559 tons and 19886 ton under two scenarios respectively, which means a 15.6% reduction by the proposed method.

Typical Scenario Analysis

To show more details of the low-carbon emission dispatch, typical scenarios are selected from the test set. First, the high power output by renewables is examined to see how the dispatch policy consumes redundant electricity. Second, during low power output by renewables, the policy has to be checked if the power supply can be stable. Third, different power flow routes are compared due to the transmission cost.

1) High renewable power output

Intuitively, during the period of high renewable power generation, the best strategy is to decrease the output from distributed generators and charge the storage unit with residual electricity that cannot be consumed by the load. In this case, test results verified the correctness of this strategy which is discovered by the agent in extensive exploration. As is shown in **Figure 5**, the output power of the wind farm at Bus six increases from 1.2 to 2 MW gradually. In the meantime, storage started charging at time step 4, and the output power of distributed generator at Bus 26 decreased from 3.4 to 2.15 MW. To maximize the usage of zero-carbon electricity generated by the wind farm, the agent decreases the output power of the distributed thermal generator and charges the storage unit. The agent makes the appropriate decision to handle the abrupt increase of output power from wind farms from both aspects: real-time power balance and low carbon emission.

Take a closer look at the phenomenon that the storage unit did not start charging until the time step 4. One reasonable explanation is that the design of the reward did not consider the carbon emission effect of the storage unit, while the distribution generator is taken into consideration. Consequently, the distributed generator has priority over the storage unit in this case.

2) Low renewable power output

Since renewable generation is heavily dependent on weather conditions, gentle wind or a large cloud could an obvious decrease in the power output. During this period, the power produced by renewables can be fully consumed, while the main issue becomes meeting the electricity demand. Typical actions are lowering the responsive load, increasing the output of distributed generators,

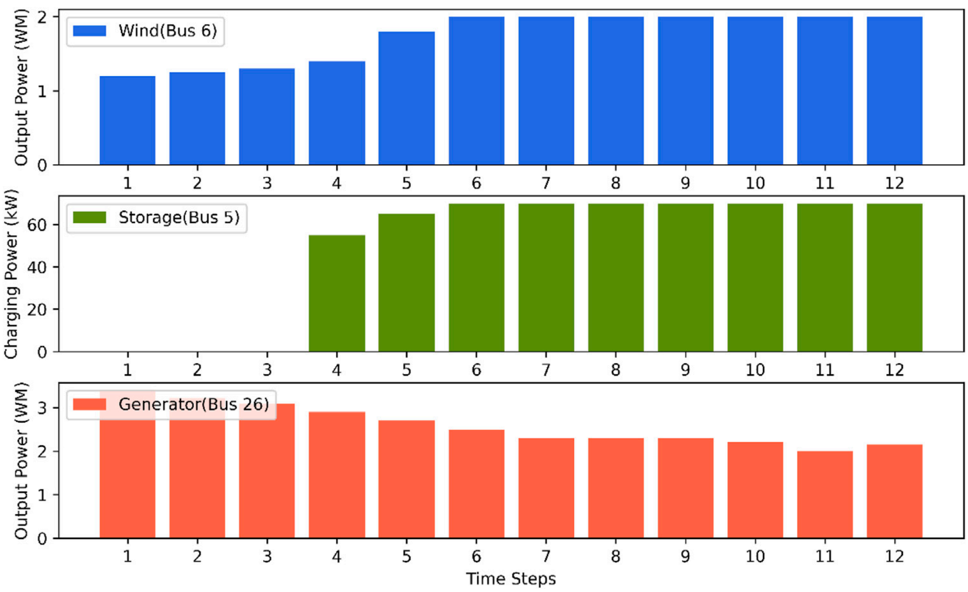

FIGURE 5 | Dispatch actions during high renewable power output.

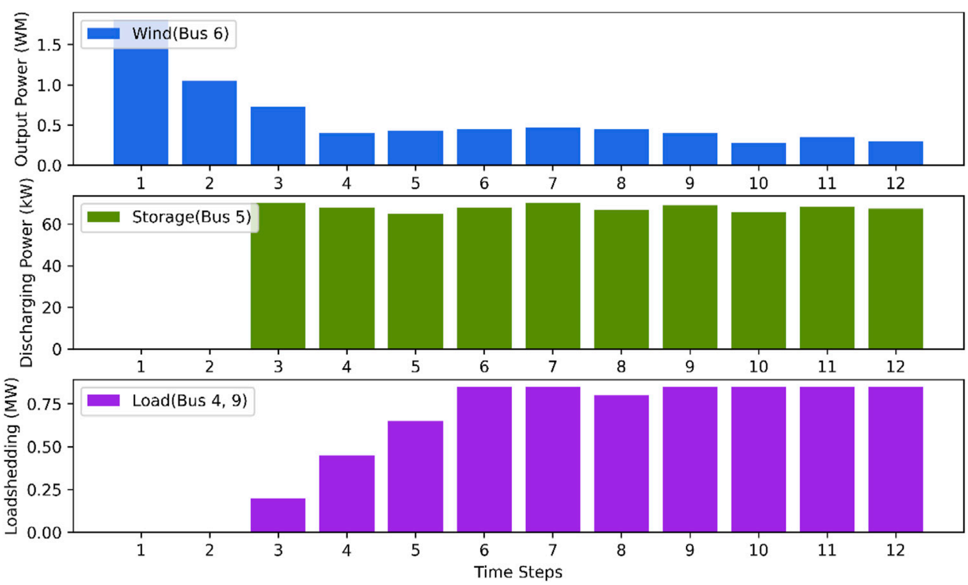

FIGURE 6 | Dispatch actions during low renewable power output.

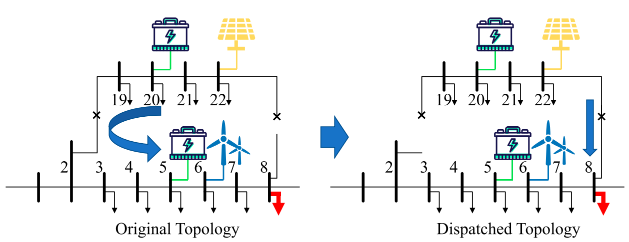

FIGURE 7 | Dispatch actions to reduce transmission cost.

and discharging the electricity storage unit. However, these actions might violate the safety constraints, so the IPO agent should make a low carbon emission and safe decision. As shown in **Figure 6**, the output power of the wind farm at Bus 6 decreases from 1.81 to 0.3 MW. To fill the power supply gap, a storage unit and responsive load are dispatched by the agent. The storage unit starts discharging at step 3 almost at the maximum output power of 70kW. Loadshedding at bus four and nine are summed up in this figure, where approximately 0.75 MW load are disconnected from the power grid.

3) Transmission cost comparison

In a distributed network, power loss during electricity transmission is between 2 and 5%. If the power flow does not in a reasonable pattern, transmission loss would go up. Moreover, the voltage of certain buses could breach the limit due to heavily loaded lines or insufficient reactive power. In this case, switches in the distribution network could come into effect by reconfiguring the topology of the grid, which improves the power flow route. In the test system, four switches can be controlled. However, these switches cannot be controlled independently due to the safety constraint on the isolated grid. It can be easily inferred from **Figure 2** that switch 2–19 and switch 8–22 cannot be disconnected simultaneously; switch 6–26 and switch 11–29 cannot be disconnected at the same time. Since this constraint cannot be violated and is difficult to depict using the mathematical expression, it is not written in the cost function and can be checked separately in the dispatch with the mentioned logical judgment. In **Figure 7**, part of the test system is a plot to compare the impact of different topologies on carbon emission. In this case, Bus eight is heavy-loaded. In an original grid, switch 3–19 is closed while switch 8–22 is open, residual power generated by the solar farm at Bus 22 has to take a long way to supply the load at Bus 8, leading to extra power loss. The agent gives dispatch orders to switches so that Bus eight and Bus 22 can be connected directly, which enables the electricity from the solar farm to be consumed in a low-carbon manner. Comparing the transmission loss of the circle and the straight route, the transmission loss is reduced by 2.51% and the corresponding carbon emission reduction is 0.12 tons for an hour.

CONCLUSION

In this study, an innovative dispatch policy is proposed to lower the carbon emission in distribution networks with proliferated renewables. As a safe RL algorithm, IPO has taken the safety constraints of the power grid into consideration, which ensures the safety of the distribution network when providing clean electricity to users. The proposed dispatch policy covers both continuous and discrete actions. For the former category, distributed generators, controllable load, and electricity storage units are included. For the latter category, switches are used to change the topology of the distribution network. To verify the effectiveness of the presented method, the case study is performed in a modification system based on the IEEE 33-bus system. Numerical results have shown that the carbon emission has decreased by 37.2% during a 365-days dataset. Moreover, all safety constraints are satisfied due to the implementation of the IPO. This study has provided guide for future development of distribution networks that appropriate local dispatch policy enables DERs to become both economic and eco-friendly.

To further extend the research, a reward can be designed considering electricity market signals. Economic profit can be an extra factor to attract users participating in local dispatching, which enlarges the dispatchable resources.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HC: writing—original draft, methodology, software, and formal analysis. YY: conceptualization, methodology, writing—review and editing, supervision, and validation. QT: investigation and writing—review and editing. YT: investigation, resources, and funding acquisition.

FUNDING

This work was funded by the Project of Shenzhen Power Supply Bureau, China Southern Power Grid (No. 090000KK52190162), 2021 Jiangsu Shuangchuang (Mass Innovation and Entrepreneurship) Talent Program, China (No. JSSCBS20210137), and the Fundamental Research Funds for the Central Universities (No. 2242022k30038).

REFERENCES

Abd El-Kareem, A. H., Abd Elhameed, M., and Elkholy, M. M. (2021). Effective Damping of Local Low Frequency Oscillations in Power Systems Integrated with Bulk PV Generation[J]. *Prot. Control Mod. Power Syst.* 6 (1), 1–13. doi:10.1186/s41601-021-00219-6

Ahmed, A., Ge, T., Peng, J., Yan, W.-C., Tee, B. T., and You, S. (2022). Assessment of the Renewable Energy Generation towards Net-Zero Energy Buildings: A Review. *Energy Build.* 256, 111755. doi:10.1016/j.enbuild.2021.111755

Al-Saffar, M., and Musilek, P. (2021). Distributed Optimization for Distribution Grids with Stochastic DER Using Multi-Agent Deep Reinforcement Learning. *IEEE Access* 9, 63059–63072. doi:10.1109/access.2021.3075247

Allan, G., Eromenko, I., Gilmartin, M., Kockar, I., and McGregor, P. (2015). The Economics of Distributed Energy Generation: A Literature Review. *Renew. Sustain. Energy Rev.* 42, 543–556. doi:10.1016/j.rser.2014.07.064

Arfeen, Z. A., Khairuddin, A. B., Larik, R. M., and Saeed, M. S. (2019). Control of Distributed Generation Systems for Microgrid Applications: A Technological Review[J]. *Int. Trans. Electr. Energy Syst.* 29 (9), e12072. doi:10.1002/2050-7038.12072

Bistline, J. E. T. (2021). Roadmaps to Net-Zero Emissions Systems: Emerging Insights and Modeling Challenges. *Joule* 5 (10), 2551–2563. doi:10.1016/j.joule.2021.09.012

Bizuayehu, A. W., Sanchez de la Nieta, A. A., Contreras, J., and Catalao, J. P. S. (2016). Impacts of Stochastic Wind Power and Storage Participation on Economic Dispatch in Distribution Systems. *IEEE Trans. Sustain. Energy* 7 (3), 1336–1345. doi:10.1109/tste.2016.2546279

Brouwer, A. S., van den Broek, M., Zappa, W., Turkenburg, W. C., and Faaij, A. (2016). Least-cost Options for Integrating Intermittent Renewables in Low-Carbon Power Systems. *Appl. Energy* 161, 48–74. doi:10.1016/j.apenergy.2015.09.090

Cao, D., Hu, W., Xu, X., Wu, Q., Huang, Q., Chen, Z., et al. (2021). Deep Reinforcement Learning Based Approach for Optimal Power Flow of

Distribution Networks Embedded with Renewable Energy and Storage Devices. *J. Mod. Power Syst. Clean Energy* 9 (5), 1101–1110. doi:10.35833/mpce.2020.000557

Caramanis, M., Ntakou, E., Hogan, W. W., Chakrabortty, A., and Schoene, J. (2016). Co-Optimization of Power and Reserves in Dynamic T&D Power Markets with Nondispatchable Renewable Generation and Distributed Energy Resources. *Proc. IEEE* 104 (4), 807–836. doi:10.1109/jproc.2016.2520758

Ghasemi, A., and Enayatzare, M. (2018). Optimal Energy Management of a Renewable-Based Isolated Microgrid with Pumped-Storage Unit and Demand Response. *Renew. Energy* 123, 460–474. doi:10.1016/j.renene.2018.02.072

Hu, J., Jiang, C., Cong, H., and He, Y. (2018). A Low-Carbon Dispatch of Power System Incorporating Active Distribution Networks Based on Locational Marginal Emission. *IEEJ Trans. Elec Electron Eng.* 13 (1), 38–46. doi:10.1002/tee.22496

Huang, Z., Xie, Z., Zhang, C., Chan, S. H., Milewski, J., Xie, Y., et al. (2019). Modeling and Multi-Objective Optimization of a Stand-Alone PV-Hydrogen-Retired EV Battery Hybrid Energy System. *Energy Convers. Manag.* 181, 80–92. doi:10.1016/j.enconman.2018.11.079

Husin, H., and Zaki, M. (2021). A Critical Review of the Integration of Renewable Energy Sources with Various Technologies[J]. *Prot. Control Mod. Power Syst.* 6 (1), 1–18. doi:10.1186/s41601-021-00181-3

IEA (2021). *Net Zero by 2050*. Paris: IEA. Available at: https://www.iea.org/reports/net-zero-by-2050.

Ju, L., Tan, Z., Li, H., Tan, Q., Yu, X., and Song, X. (2016). Multi-objective Operation Optimization and Evaluation Model for CCHP and Renewable Energy Based Hybrid Energy System Driven by Distributed Energy Resources in China. *Energy* 111, 322–340. doi:10.1016/j.energy.2016.05.085

Li, Y., Hao, G., Liu, Y., Yu, Y., Ni, Z., and Zhao, Y. (2021). Many-objective Distribution Network Reconfiguration via Deep Reinforcement Learning Assisted Optimization Algorithm[J]. *IEEE Trans. Power Deliv.* 37, 2230. doi:10.1109/TPWRD.2021.3107534

Lin, C., Wu, W., Chen, X., and Zheng, W. (2017). Decentralized Dynamic Economic Dispatch for Integrated Transmission and Active Distribution Networks Using Multi-Parametric Programming[J]. *IEEE Trans. Smart Grid* 9 (5), 4983–4993.10.1109/TSG.2017.2676772

Liu, Y., Ding, J., and Liu, X. (2020). IPO: Interior-Point Policy Optimization under Constraints. *AAAI* 34 (04), 4940–4947. doi:10.1609/aaai.v34i04.5932

Mohammadjafari, M., Ebrahimi, R., and Darabad, V. P. (2020). Multi-objective Dynamic Economic Emission Dispatch of Microgrid Using Novel Efficient Demand Response and Zero Energy Balance Approach[J]. *Int. J. Renew. Energy Res.* 10 (1), 117–130. doi:10.20508/ijrer.v10i1.10322.g7846

Nguyen, D. H. (2020). Optimal Solution Analysis and Decentralized Mechanisms for Peer-To-Peer Energy Markets[J]. *IEEE Trans. Power Syst.* 36 (2), 1470–1481. doi:10.1109/TPWRS.2020.3021474

Ou, Y., Roney, C., Alsalam, J., Calvin, K., Creason, J., Edmonds, J., et al. (2021). Deep Mitigation of CO2 and non-CO2 Greenhouse Gases toward 1.5° C and 2° C Futures[J]. *Nat. Commun.* 12 (1), 1–9. doi:10.1038/s41467-021-26509-z

Pehl, M., Arvesen, A., Humpenöder, F., Popp, A., Hertwich, E. G., and Luderer, G. (2017). Understanding Future Emissions from Low-Carbon Power Systems by Integration of Life-Cycle Assessment and Integrated Energy Modelling. *Nat. Energy* 2 (12), 939–945. doi:10.1038/s41560-017-0032-9

Wang, D., Qiu, J., Reedman, L., Meng, K., and Lai, L. L. (2018). Two-stage Energy Management for Networked Microgrids with High Renewable Penetration. *Appl. Energy* 226, 39–48. doi:10.1016/j.apenergy.2018.05.112

Wang, Y., Huang, Z., Shahidehpour, M., Lai, L. L., Wang, Z., and Zhu, Q. (2019). Reconfigurable Distribution Network for Managing Transactive Energy in a Multi-Microgrid System[J]. *IEEE Trans. smart grid* 11 (2), 1286–1295. doi:10.1109/TSG.2019.2935565

Ye, Y., Qiu, D., Sun, M., Papadaskalopoulos, D., and Strbac, G. (2019). Deep Reinforcement Learning for Strategic Bidding in Electricity Markets[J]. *IEEE Trans. Smart Grid* 11 (2), 1343–1355. doi:10.1109/TSG.2019.2936142

Zhang, Y., Ai, X., Wen, J., Fang, J., and He, H. (2018). Data-adaptive Robust Optimization Method for the Economic Dispatch of Active Distribution Networks[J]. *IEEE Trans. Smart Grid* 10 (4), 3791–3800. doi:10.1109/TSG.2018.2834952

Zhang, X., Liu, Y., Duan, J., Qiu, G., Liu, T., and Liu, J. (2021). DDPG-based Multi-Agent Framework for SVC Tuning in Urban Power Grid with Renewable Energy Resources. *IEEE Trans. Power Syst.* 36 (6), 5465–5475. doi:10.1109/tpwrs.2021.3081159

Zhou, X., Ai, Q., and Yousif, M. (2019). Two Kinds of Decentralized Robust Economic Dispatch Framework Combined Distribution Network and Multi-Microgrids. *Appl. Energy* 253, 113588. doi:10.1016/j.apenergy.2019.113588

Conflict of Interest: Author QT is employed by the Shenzhen Power Supply Bureau.

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.

The handling editor HW is currently organizing a Research Topic with the author(s) YY.