Article

Fault Recovery Methods for a Converged System Comprised of Power Grids, Transportation Networks and Information Networks

Geng Zhang [1,2]**, Chenxu Liu** [1]**, Hao Jiang** [1,*] **and Jiye Wang** [3]

[1] Electronic Information School, Wuhan University, Wuhan 430072, China; zhanggeng@epri.sgcc.com.cn (G.Z.); liuchenxu@whu.edu.cn (C.L.)
[2] China Electric Power Research Institute, Beijing 100086, China
[3] State Grid Digital Technology Holdings Co., Ltd., Beijing 100086, China; wangjiye@epri.sgcc.com.cn
[*] Correspondence: jh@whu.edu.cn

Abstract: Recently, triple-network convergence systems (TNCS) have emerged from the deep integration of the power grid, transportation networks, and information networks. Fault recovery research in the TNCS is important since this system's complexity and interactivity can expand the fault's scale and increase the fault's impact. Currently, fault recovery focuses primarily on single power grids and cyber–physical systems, but there are certain shortcomings, such as ignoring uncertainties, including generator start-up failures and the occurrence of new faults during recovery, energy supply–demand imbalances leading to system security issues, and communication delays caused by network attacks. In this study, we propose a recovery method based on the improved twin-delayed deep deterministic algorithm (TD3), factoring in the shortcomings of the existing research. Specifically, we establish a TNCS model to analyze interaction mechanisms and design a state matrix to represent the uncertainty changes in the TNCS, a negative reward to reflect the impact of unit start-up failures, a special reward to reflect the impact of communication delay, and an improved actor network update mechanism. Experimental results show that our method obtains the optimal recovery decisions, maximizes restoration benefits in power grid failure scenarios, and demonstrates a strong resilience against communication delays caused by DoS attacks.

Keywords: triple-network convergence; fault recovery; TD3 algorithm; DoS attack; resilience

1. Introduction

In recent years, the TNCS has emerged due to the rapid development and deep integration of power grids, transportation networks, and information networks caused by the widespread adoption of electric vehicles. This convergence system enables efficient coordination among energy flow, traffic flow, and information flow [1,2]. However, the increasing complexity and interdependence of the system have also created favorable conditions for fault propagation, in which local disturbances can easily initiate cascading failures across networks, resulting in severe system risks such as wide-scale power outages [3–7]; therefore, research on fault recovery method for the TNCS is essential.

The development of a recovery method for the TNCS necessitates a clear understanding of its interaction mechanisms. However, there is a paucity of explicit domestic and international research on modeling the interaction mechanisms of such systems, which adds to the complexity of the study. Furthermore, there is a paucity of research on fault recovery methods related to the TNCS. Despite this, a body of research exists on fault recovery for power grids and power cyber–physical systems. In reference [8], for instance, a cascading fault recovery model based on the sequential recovery graph (SRG) was suggested for power transmission networks and was proven by simulations. In reference [9], an optimization model for the entire black start restoration process in gearbox systems was

formulated and solved using a linearized hybrid integer linear programming model and the L-shaped algorithm, taking into consideration the uncertainty of wind power output. Reference [10] analyzed the potential delays caused by information system failures in power system generation, transmission line operation, and load restoration. In these studies, however, the impact of generator start-up failures on restoration decisions was not considered. Reference [11] proposed a parallel recovery method for power systems. Although this method considered uncertainties during the restoration process, it did not consider the possibility of new faults occurring in power systems during component restoration.

The recovery of power system components can result in alterations to the power system's topology, power flow, and power levels, which may cause problems such as line overloads, imbalances in energy supply and demand in certain regions, and system instability. However, existing research lacks studies on the risks of new faults or system instability that may arise from component recovery. For instance, in reference [12], a two-stage recovery strategy was proposed for power facility restoration after hurricanes, focusing on the design of distributed generation and maintenance personnel scheduling but excluding the analysis of new faults that might occur from power system component recovery. References [13–15] also ignored the occurrence of new faults.

It is evident that existing research typically establishes optimization models for different fault scenarios and solves them using various optimization algorithms; however, they lack consideration of uncertainties that may occur during the actual restoration process, such as generator start-up failures and faults triggered by recovery actions, and their impact on recovery decisions. Consequently, these uncertainties should be considered when studying fault recovery methods for TNCS. Such research methods would be more applicable to real-world circumstances.

Moreover, due to the high complexity and interactivity of the TNCS, network attacks can exploit intricate network interaction mechanisms and exacerbate the hazards of faults with relative ease. By manipulating power grid data and injecting false information, for instance, the assessment of the system's operational status may deviate, resulting in erroneous decisions by administrators and potentially triggering large-scale cascading failures [16–19]. By injecting a large number of useless requests and obstructing communication channels, DoS attacks can cause an increase in communication delays and, thus, initiate or exacerbate power outages [20–24]. An example of this was the DoS attack launched by a hacker group against a power company in the western United States in 2019, which resulted in communication disruption between the control center and various site devices [25]. There is limited research on restoration decisions for power cyber–physical systems that take network attacks into account at present. While the optimization strategy proposed in [26] considered the impact of information system failures caused by DoS attacks on the restoration process, the solving algorithm lacked the flexibility to handle uncertainty factors and became less efficient as the system scale increased.

The TD3 algorithm, from the discipline of deep reinforcement learning, can be used to address this issue [27]. By appropriately designing the state space and reward function, uncertainties can be accounted for, and the neural network can effectively address the curse of dimensionality problems caused by the large scale of the TNCS [28,29]. In addition, compared to the deep Q-network (DQN), actor–critic (AC), and deep deterministic policy gradient (DDPG) algorithms [30–32], the TD3 algorithm exhibits a better capability to suppress network overestimation and provides more stable network training. Consequently, the TD3 algorithm demonstrates significant advantages over existing methods in addressing the complexity of the system's scale, uncertainties, and stability during the solution process, making it more suitable for solving the model. However, the optimization objective of the TD3 algorithm is to find actions that maximize the action-value function in different states without considering the specific role and impact of recovery actions in particular scenarios, such as system security; therefore, improvements to the TD3 algorithm are necessary to ensure system security.

Based on the above analysis, this study makes the following main contributions to the research on fault recovery methods in TNCS:

1. Focusing on the charging behavior of electric vehicles, a TNCS model is established to reveal the underlying interaction mechanisms;
2. An efficient fault recovery method for TNCS is proposed, incorporating an improved TD3 algorithm and considering communication delays. By designing and improving the TD3 algorithm, the uncertainties and security issues in the restoration process are considered, leading to the design of an effective recovery algorithm. In addition, the resilience of the algorithm is evaluated by introducing DoS attacks in the context of power grid faults. Lastly, the efficacy of the proposed recovery method is demonstrated through simulation experiments.

2. Methods

2.1. Modeling

The TNCS system comprises three integral components: the power grid, the information network, and the transportation network. Initially, we model the power grid, the scale of which is contingent upon actual demands. Subsequent to this foundation, we construct corresponding transportation and information network models, enabling the circulation of energy and information flows among the three components.

2.1.1. Overall Framework

We represent the overall framework of the TNCS system as a whole, consisting of three layers, as shown in Figure 1. The execution layer consists of the actual power grid and transportation network, which is responsible for the reliable operation of energy interactions. The coupling layer mainly consists of power collection and control devices, as well as communication channels, and is responsible for information acquisition, execution of control commands, and information data transmission. The control layer, primarily based on the information network, is responsible for state monitoring and dispatch control.
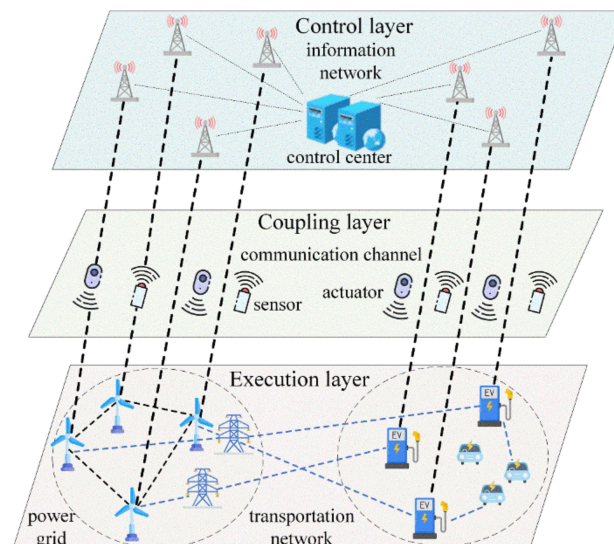
Figure 1. The TNCS model. In the power grid, wind turbines are considered as power nodes, supplying energy to charging stations in the transportation network via blue electrical lines. The charging stations, in turn, deliver energy to electric vehicles through blue lines. Operational data from both the turbines and charging stations are relayed to various information nodes in the control layer via thicker black transmission links. Within the information network, we use the base station icon to represent the information nodes, fulfilling the roles of data reception and transmission. Furthermore, while the control center is typically an amalgamation of multiple devices and systems, we employ a computer icon as its symbol for representational simplicity.

Within the execution layer, the power grid continuously powers the transportation network through electrical transmission lines. Their operational parameters are collected by sensor devices in the coupling layer and relayed to the information nodes within the control layer. These nodes aggregate the gathered data into the control center. Subsequently, the directives generated by the control center are conveyed to the physical systems in the execution layer through actuators within the coupling layer.

2.1.2. Execution Layer

The power grid topology is a directed graph, $G_P = (V_P, E_P)$, where V_P represents power nodes, including generators, loads, and circuit breakers, and E_P represents the transmission lines. The topological information for the power grid is represented by a matrix T^P whose elements T_{ij}^P adhere to the following rules: $T_{ij}^P = 1$ denotes the presence of a line between nodes i and j, while $T_{ij}^P = 0$ denotes the absence of a line. The nodal admittance matrix B, nodal injection power vector P, line power flow vector f, and nodal phase angle vector θ are established. Based on the DC power flow model, the power flow information matrix PF of the power grid can be obtained from the following equations:

$$PF = diag(P) + f \tag{1}$$

$$B\theta = P \tag{2}$$

$$\left(\theta Q - Q^T \theta^T\right) \circ B = f, \tag{3}$$

where the value of Q is $[1, 1, 1, ..., 1]$ and $diag(P)$ represents a diagonal matrix with vector P as its diagonal elements.

The road topology of the transportation network is an undirected graph, $G_T = (V_T, E_T)$, where V_T represents road nodes (including charging stations) and E_T represents paths between them. A matrix A_T is established to represent the topological information of the transportation network, with matrix elements A_{ij}^T following the rules: $A_{ij}^T = 1$ indicates the presence of a connected road between road nodes i and j, while $A_{ij}^T = 0$ indicates the absence of a connected road. A decision vector SE is used to assess whether each road node has a charging station, where $SE_i = 1$ indicates the presence of a charging station in road node i and $SE_i = 0$ indicates its absence. A charging station load vector P^c is introduced for the transportation network, where P_i^c represents the load at the charging station located on road node i and can be obtained from the following equations:

$$P_i^{c\prime} = \sum_{j=1}^{N_i^S} CSO_{ij} \circ CSW_{ij} \circ CSP_{ij} \tag{4}$$

$$P_i^c = SE_i \circ P_i^{c\prime}, \tag{5}$$

where N_i^S represents the number of charging piles in the charging station located at road node i. CSO_{ij} indicates the operational status of the j-th charging pile in the charging station at road node i, with a value of 1 indicating that it is operational and 0 indicating that it is not. CSW_{ij} represents the operational status of the j-th charging pile in the charging station at road node i, with 1 indicating operation and 0 indicating inactivity. CSP_{ij} represents the charging power of the j-th charging pile at the road node i charging station.

To ensure the safety of charging services, the maximum charging power for each charging pile must be specified. Establishing a maximum power vector CSM, where CSM_{ij} represents the maximum charging power of the j-th charging pile in the charging station at road node i.

2.1.3. Coupling Layer

The matrices S^P and S^T are respectively defined to represent sensor deployment in the power grid and transportation network. In S^P, diagonal elements denote sensors on the nodes of the power grid, while off-diagonal elements represent sensors on the power lines. In S^T, the diagonal elements represent sensors on the road node charging stations. The elements of both matrices are assigned the value 1 to denote the presence of a sensor and the value 0 to denote the absence of a sensor. The actuator matrices A^P and A^T are defined in the same way as the sensor matrices.

The uplink data communication channel matrix contains UC^P for the power grid and UC^T for the transportation network. The matrix elements, UC_{ij}, satisfy $UC_{ij} \in \{0, 1\}$. In UC^P, $UC_{ii}^P = 1$ indicates the presence of an uplink data communication channel between the sensor on the power grid node and the information network control center, transmitting node power information. $UC_{ij}^P = 1$ indicates the existence of an uplink data communication channel between sensors on power grid lines i-j and the information network control center, transmitting current flow information and circuit breaker status. In UC^T, $UC_i^T = 1$ denotes the presence of an uplink data communication channel between the sensors in the charging station at road node i and the information network control center, transmitting charging station operational information. The downlink data communication channel matrix is defined in the same way as UC^P and UC^T, with different transmitted content. It carries commands for the charging station's circuit breaker open or close, node power adjustment, charging pile open or close, and maximal charging power adjustment.

2.1.4. Control Layer

The information network topology is defined as a directed graph, $G_C = (V_C, E_C)$, where V_C represents the information nodes and E_C represents the communication links between nodes. The connectivity status of the communication links is represented by the link status matrix C, with matrix elements $C_{ij}(i \neq j) = \{0, 1\}$. A value of $C_{ij} = 1$ indicates that an effective connection has been established between information nodes i and j. The control center in the information network adjusts the operational status of the power grid and transportation network based on the received data. The storage of received information includes PM^f and PM^t for power flow information (PF) and power grid topology information (T^P), respectively, while TM^S represents the storage of received charging station information. Utilizing this information, the reduction in output of generator nodes and the reduction in load of load nodes are calculated, thereby calculating the decrease in power at the charging stations in the transportation network, as illustrated below:

$$\sum_i^N P_i^a \sum_{j=1}^{N_i^S} CSO_{ij}' = \sum_{i=1}^N P_i^C - P_i^S \tag{6}$$

$$\left| P^C - P^S \right| \leq |LS|, \tag{7}$$

where LS satisfies Equations (1)–(3), and is used to represent the reduction in output of power nodes or the reduction in load of load nodes. P^S is the drop value in charging power at road node i, while P_i^a is the adjusted average charging power of the charging piles in the charging stations. CSO_{ij}' refers to the adjusted operational state of the j-th charging pile in the i-th road node. N represents the number of road nodes in the transportation network. We assume that the charging stations will bear the maximum extent of power grid load variations.

The control center then generates and distributes decision instructions to the executive layer.

2.2. Fault Recovery Method

Common power grid failures often arise from localized disturbances or anomalies, subsequently spreading and affecting the operational status of other lines or nodes, potentially leading to a system-wide collapse, which is referred to as cascading failures. Cascading failures have a more pronounced impact and are more frequently encountered in real-life scenarios; consequently, our design of fault recovery methods primarily targets multi-fault situations.

2.2.1. Design of Restoration Model

In this section, we define important indicators for power grid faults in the context of the TNCS, assessing the importance of faulty power lines. With the objective of maximizing restoration benefits, a restoration model is established.

We assume that power grid faults in the TNCS result from the overload-induced disconnection of power lines, leading to the redistribution of power flow, reduction in output of generator nodes, and load shedding at load nodes. Concurrently, adjustments are made to the charging power of the charging stations in the transportation network corresponding to the affected power nodes, which may impact the travel of EV users. The importance of power lines, η, is determined by the importance of its connected power nodes (η_P) and the importance of the charging stations on the corresponding coupled road nodes (η_S), as shown below.

$$\eta = \lambda_1 \cdot \omega \eta_P + \lambda_2 \cdot (1 - \omega) \eta_S, \eta > 0, \eta_P > 0, \eta_S > 0, 0 < \omega < 1 \tag{8}$$

where ω represents the weight coefficient. λ_1 and λ_2 are binary variables that indicate the existence of the respective item, taking a value of 1 if it exists and 0 otherwise. η_P consists of the degree, generation capacity, and load of the power nodes and η_S consists of the degree of the road nodes and the number of electric vehicles affected, as shown below:

$$\eta_P = \lambda_3 \frac{P_g}{\sum_{i \in gen} P_i^g} + \lambda_4 \frac{\gamma_d}{\sum_{i \in |V_P|} \gamma_i} + \lambda_5 \frac{P_l}{\sum_{i \in load} P_i^l} \tag{9}$$

$$\eta_S = \lambda_6 \frac{\gamma_t}{\sum_{i \in |V_T|} \gamma_i} + \lambda_7 \frac{\chi_t}{\sum_{i \in |V_T|} \chi_i^t} \tag{10}$$

$$\chi_t = \chi_{t1} + \chi_{t2} \tag{11}$$

$$\chi_{t1} = \chi_{t1} + 1, \sum_{i=1}^{K} \sum_{j=1}^{|V_T|} \sum_{k=1}^{N^S} \frac{IF\left(\left|TS_i^p(j,k)\right| > 0\right.}{AND\ CSN_i(j,k) \neq 0, 1, 0\right)} \tag{12}$$

$$\chi_{t2} = \chi_{t2} + 1, \sum_{i=1}^{K} \sum_{j=1}^{|V_T|} \sum_{k=1}^{N^S} \frac{IF\left(\left|TS_i^p(j,k)\right| > 0\right.}{AND\ CSQ_i(j,k) \neq 0, 1, 0\right)} \tag{13}$$

where λ_3, λ_4, λ_5, λ_6, and λ_7 are binary variables indicating the existence of the respective items. P_g represents the power output of generator nodes and $\sum_{i \in gen} P_i^g$ represents the total power output of generator nodes. γ_d represents the degree of nodes in the power grid, while $\sum_{i \in |V_P|} \gamma_i$ represents the sum of the degree of power nodes. γ_t represents the degree of road nodes in the transportation network, while $\sum_{i \in |V_T|} \gamma_i^t$ is the sum of the degree of road nodes. P_l represents the power load of the load nodes in the power grid, while $\sum_{i \in load}^m P_i^l$ denotes the total power load of the load nodes in the power grid. χ_t obtained by Equations (11)–(13) represents the number of EVs impacted during travel, while $\sum_{i \in |V_T|} \chi_i^t$ represents the total number of EVs impacted during travel. Equation (12) calculates the adjusted number of charging station users χ_{t1}, while Equation (13) calculates the adjusted number of charging station electric vehicles χ_{t2}. TS_i^p represents the power

adjustment matrix for the i-th charging station, CSN_i represents the identifier of the current charging station users, and CSQ_i represents the identifier of the current charging station reservation users.

After calculating the importance of each power line, they are sorted in descending order to determine the restoration priority. Γ is defined as the quantified priority value, and e_p represents the number of restoration steps. The optimization objective is to maximize the restoration benefit, which consists of the recovery power from the power grid and transportation network, based on the importance of the power lines.

$$max\sum_{m}^{e_p}\Gamma\left(LS_{rm} + P_{rm}^{S}\right)\tag{14}$$

$$s.t.\Gamma = 1 - 0.02\left(e_p - 1\right), e_p \geq 1\tag{15}$$

$$min(LS) \leq LS_{rm} \leq max(LS)\tag{16}$$

$$min\left(P^S\right) \leq P_{rm}^{S} \leq max\left(P^S\right)\tag{17}$$

$$P_{min}^{G} \leq P^G \leq P_{max}^{G}\tag{18}$$

$$V_{min} \leq V \leq V_{max}\tag{19}$$

$$f_{min} < f < f_{max}\tag{20}$$

$$0 < T^{up} < T^h\tag{21}$$

$$\sum_{i \in gen} P_i^g = \sum_{i \in load} P_i^l\tag{22}$$

where $\Gamma\left(LS_{rm} + P_{rm}^{S}\right)$ represents the restoration benefit obtained from restoring the fault at step m. P^G is the active power of the generator, and V is the voltage at the terminals of the power system. Equation (20) represents the line flow constraint, where power flow varies with the restoration of the faults. Exceeding the limits of line flow, f_{min} or f_{max}, may result in overload issues, leading to the disconnection or damage of the concerned line or other lines. In Equation (21), T^{up} represents the unit start-up time, while T^h represents the maximum start-up time of the unit, ensuring the required time for thermal start-up. Equation (22) represents the energy supply–demand balance constraint in the power grid, ensuring the balance between actual supply and demand. Imbalances can lead to voltage fluctuations, frequency deviations, and overloaded operation, causing power equipment failures or system security issues such as a power grid collapse. In this study, we assume that energy imbalance can cause line overloads and disconnection so that the fault recovery should avoid energy supply–demand imbalances.

2.2.2. Design of Restoration Algorithm

In the context of fault recovery in the power grid of TNCS, the restoration sequence of each faulty line significantly influences the restoration benefit, giving rise to a decision-making problem. In this section, we focus on the design and improvement of the TD3 algorithm in the field of deep reinforcement learning with the aim of determining the optimal restoration strategy. Figure 2 illustrates the restoration algorithm structure based on the TNCS.
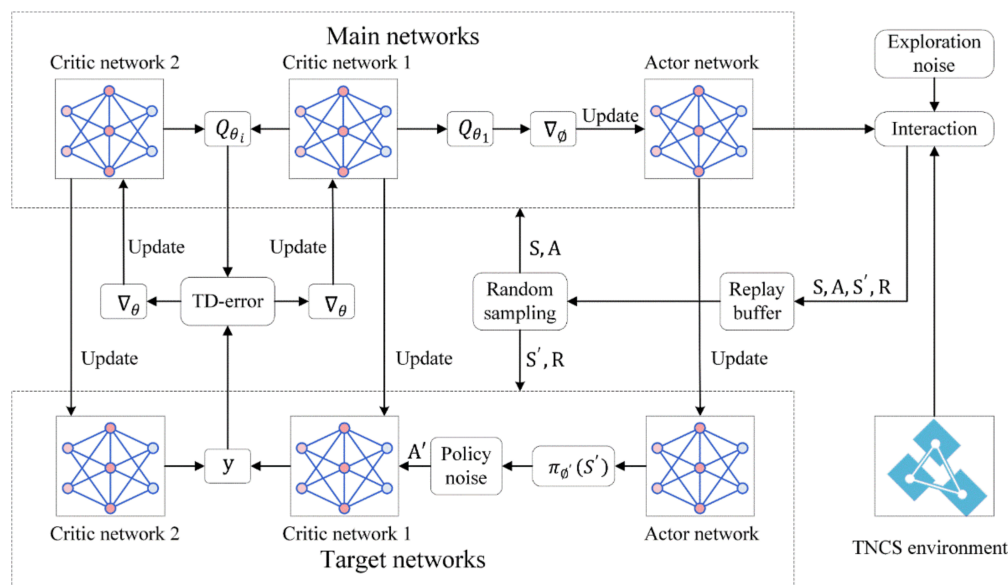
Figure 2. Structure of the restoration algorithm. First, the TNCS environment interacts with the actor network in the main networks, incorporating exploration noise. Specifically, this interaction indicates the state matrix S, as the input to the actor network, and the actor network chooses an action based on S. Then, the action can change the state matrix, resulting in a certain reward. Second, the obtained transitions from the interaction (transitions primarily include S, A, S', and R) are preserved in the replay buffer via experience replay. To train and update the actor network and critic network in the main networks, a batch-sized transition sample is selected at random. During the update of the critic network, a TD-error is generated and gradient descent is performed while introducing policy noise into the process. The actor network is updated using gradient ascent to maximize Q_{θ_1}. The Target networks are updated after a certain number of updates to the main networks.

Specifically, we consider the control center as the agent and define the action space as the set of all available restoration actions from which the control center can choose in the TNCS environment, denoted as Ac, and action A belongs to Ac. In the restoration model, the recovery of faults can cause new faults. To address this uncertainty, we employ a state matrix S to represent the state space, reflecting the changes in the state of the TNCS. As shown in Figure 3, each row of the state matrix represents a power node or line, and each column represents a fault type. The arrangement follows the order of power node and line numbering, as well as fault type. Based on the rows and columns, the location and type of faults can be determined. An element value of 1 indicates no fault, while 0 indicates that a fault exists. Consequently, the changes in the state of the TNCS caused by fault recovery and occurrence can be reflected through the variation of element values.
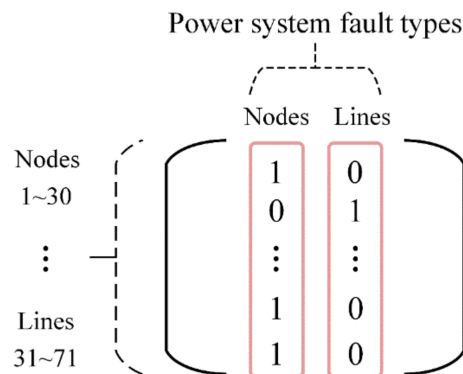
Figure 3. State matrix. We assume the presence of 30 power nodes, 41 lines, and fault types including faults at power nodes and lines.

The design of the reward function is as follows:

$$R = \begin{cases} I_d + I_b + I_m, (\mu = 1, \varepsilon = 1) \\ I_d + I_c, (\mu = 0) \\ I_d - 1, (\mu = 1, \varepsilon = 0) \end{cases} \tag{23}$$

$$I_m = \frac{LS_{rm} + P_{rm}^S}{max(LS) + max(P^S)} \tag{24}$$

$$I_c = \left(1 + \beta \frac{h_t}{H}\right)\Gamma + \alpha \frac{e_c}{E}. \tag{25}$$

The control center conducts a status check on the TNCS at the beginning of each restoration step, and when the check is complete, the reward I_d is obtained. If no data information is received from a node or line in the execution layer within time T_l, it is considered communication delays, $\mu = 0$ is set, and the reward I_c is obtained. Equation (25) represents the impact of communication delays on restoration decisions, where β and α are proportional coefficients with respective ranges of [0,1] and (0,1). H represents the total number of restoration stages, indicating the duration of the entire restoration process, while h_t represents the restoration stage at which the communication delay occurs. The restoration stage E represents the maximum number of restoration steps required to address communication faults (with a single stage consisting of multiple consecutive restoration steps) and e_c represents the actual number of restoration steps required for communication fault recovery. $\mu = 1$ denotes normal communication, with $\varepsilon = 1$ indicating the effectiveness of the restoration action, resulting in the reward I_b, and the restoration benefit I_m calculated from Equation (24). When ε equals zero, the restoration action is ineffective, resulting in a negative reward of -1.

The effectiveness of a restoration action refers to its objective of targeting the faulty lines without causing any additional faults. Restoring a power node requires at least one connected power line to be operational (excluding black-start nodes). For a power generation node, Equation (21) must be satisfied, and if not, it is considered a failed unit start-up, which is uncertain and may be caused by unpredictable factors. In the case of a failed unit start-up, the reward function returns -1, reflecting the impact of this uncertainty on restoration decisions, and we can enhance the practicality of our method by employing this approach.

In response to potential system security issues caused by an imbalance between energy supply and demand during the restoration process, the TD3 algorithm is improved as follows:

$$\rho = \lambda_8 \eta \tag{26}$$

$$q = Q_{\theta_1}(S, A)\big|_{A=\pi_\phi(S)} \cdot (1 - \rho_a). \tag{27}$$

Equation (27) adds a multiplicative term $(1 - \rho_a)$ in the original estimated value $Q_{\theta_1}(S, A)\big|_{A=\pi_\varnothing(S)}$ used for updating the actor network, where q represents the improved estimated value and ρ_a belongs to the set ρ. ρ represents the set of line importance, as determined by Equation (26). λ_8 is a binary variable with values of 0 and 1, and it is set to 1 if the restoration action causes an energy supply–demand imbalance in the system and 0 otherwise.

The addition of the multiplicative term can affect the original estimated value after the improvement. When high-priority lines are disconnected due to an energy supply–demand imbalance, the multiplicative term becomes smaller, resulting in a decrease in estimated value and indicating that the restoration actions under the current state pose a security risk to the system. Therefore, the actor network can conclude that the chosen restoration action is not optimal.

Lastly, the neural networks used in the algorithm consist of an input layer, fully connected hidden layers, and an output layer. The specific parameter setting depends on the simulation experimental scenarios.

3. Experiments and Results

3.1. Experimental Settings

Based on the TNCS model, a simulation experimental scenario is constructed as follows.

The power grid uses the IEEE 30-bus power system [33], as shown in Figure 4. The transportation network has 5000 EVs and 30 nodes with 50 charging piles at each road node. Each charging pile has a maximal output of 20 kW. The nodes of the transportation network and the power grid with the same number are coupled. The information network consists of 101 information nodes, where the first 30 nodes correspond to the power grid nodes, with node 10 serving as the control center. Nodes 31 to 71 correspond to the power lines, while the remaining nodes correspond to the transportation network nodes. The simulation experiments are implemented using MATLAB programming without GPU acceleration techniques.
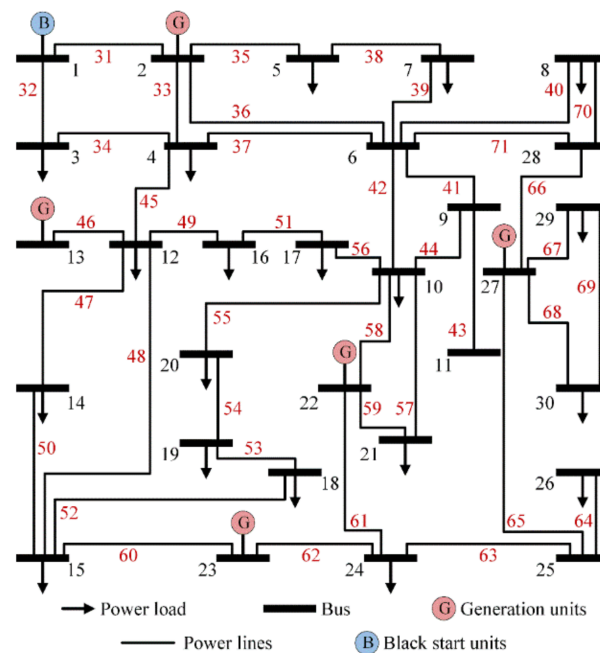
Figure 4. IEEE 30-bus power system diagram. The power lines are numbered from 31 to 71.

3.2. Small-Scale Power Grid Faults

In this section, we simulated small-scale faults in the power grid of the TNCS. Specifically, we induce faults in the power lines with line numbers 31, 32, 33, 34, 36, and 37, and these lines are placed in an open-circuit state. The specific experimental parameter settings are as follows.

The actor network consists of an input layer with 142 neurons, a first hidden layer with 256 neurons, a second hidden layer with 128 neurons, and an output layer with 6 neurons. The target actor network is updated once after every five updates of the actor network.

The critic network consists of an input layer with 142 neurons, a first hidden layer with 256 neurons, a second hidden layer with 256 neurons, and an output layer with 1 neuron. The target critic network is updated once after every five updates of the critic network.

The ReLU function is used as the activation function in the hidden layers of all networks. The actor network is updated once after every three updates of the critic network. Set the discount factor to 0.99, the exploration noise standard deviation to 0.15, the policy

noise standard deviation to 0.3, the batch size to 64, the total number of episodes to 1000, and the ω, I_d, I_b, and τ values to 0.5, 0.1, 0.1, and 0.005, respectively. The parameter τ, specifically, is used in the soft update.

The simulation experimental results are as follows.

The result in Figure 5 indicates that the actor network is well-trained and able to make accurate and effective decisions.
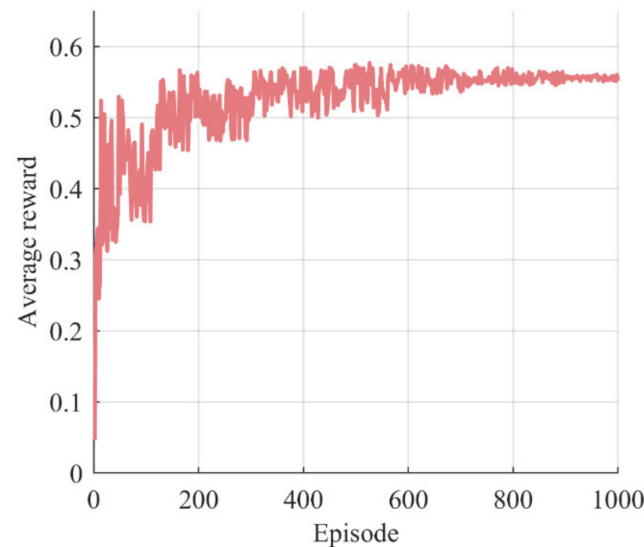
Figure 5. Learning curve. We can observe that the learning curve exhibits significant fluctuations during the first 300 or so episodes. Approximately between 300 and 700 episodes, the fluctuations decrease. After roughly 700 episodes, the average reward converges on a stable value.

The obtained restoration scheme is shown in Table 1. The restoration sequence is primarily determined by the order of line restoration; once a line is restored, the connected nodes automatically begin their restoration process (except for black-start nodes). The restoration sequence refers specifically to the order of initiating restoration. As a result, the restoration processes for multiple faults can proceed simultaneously. For example, the restoration of line 36 can be initiated during the restoration of node 1.

Table 1. Restoration sequence.

Steps	Power Grid Nodes	Power Grid Lines
1	1, 2	31
2	6	36
3	4	33
4		37
5		32
6		34

To validate the optimality of the restoration sequence, we compared it with several backup schemes (utilizing a solver for computation) based on the restoration benefit criterion. As shown in Table 2, the restoration benefit obtained by our proposed scheme, denoted F_1, was from 15.36% to 25.52% greater than that of the other schemes, indicating that our scheme is optimal in terms of restoration benefit.

Table 2. Comparison of schemes.

Scheme	Restoration Sequence	Restoration Benefit (MW)
F_1	31, 36, 33, 37, 32, 34	2.8239
F_2	33, 36, 37, 32, 31, 34	2.2689
F_3	31, 32, 34, 37, 36, 33	2.3564
F_4	36, 32, 34, 31, 33, 37	2.3901
F_5	34, 33, 36, 32, 31, 37	2.1032
F_6	31, 37, 32, 34, 36, 33	2.2946

Additionally, Figure 6 shows the improvement effect of the TD3 algorithm.
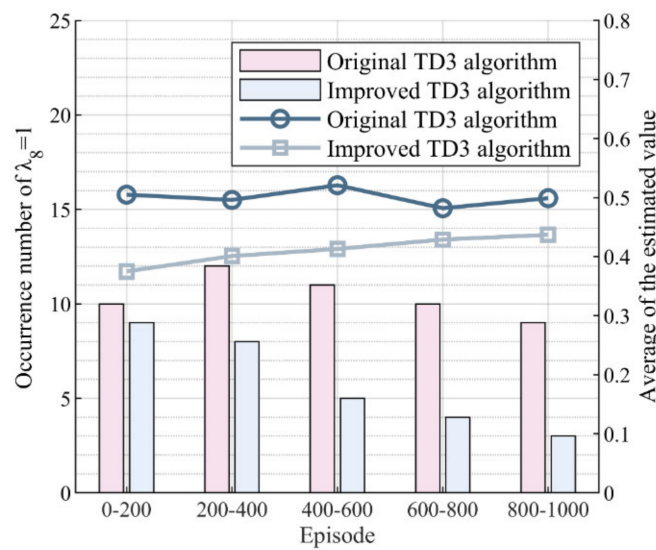
Figure 6. Improvement effect of the TD3 algorithm. The values of the bar indicate the number of occurrences of $\lambda_8 = 1$ per 200 episodes, while the values of the line indicate the average of the estimated value every 200 episodes.

It can be observed that the occurrence number of $\lambda_8 = 1$ was lower in the improved algorithm compared to the unimproved version. This indicates that the evaluation of restoration actions in the improved algorithm was not solely based on maximizing $Q(S, A)$, but also considered system security issues caused by energy supply–demand imbalances, as represented by the inclusion of a multiplicative term. Using this method, it is possible to lower the original estimated values $Q_{\theta_1}(S, A)\big|_{A=\pi_\varnothing(S)}$, as shown by the lower values of the line in the figure for the improved TD3 algorithm compared to the unimproved TD3 algorithm. When the chosen restoration action leads to an energy supply–demand imbalance, λ_8 equals 1, the estimated value decreases, and the actor network receives feedback indicating that the chosen restoration action is not optimal. As the number of episodes increases, the actor network gradually learns to avoid restoration actions that may cause energy supply–demand imbalances. Consequently, in the figure, we can observe that the frequency of $\lambda_8 = 1$ decreased over time, with only three occurrences between episodes 800 and 1000, representing a probability of 1.5%. This is a 66.67% reduction compared to the unimproved TD3 algorithm, and demonstrates the effectiveness of our improvements to the TD3 algorithm in reducing the probability of system security issues caused by energy supply–demand imbalances during the restoration process.

3.3. Large-Scale Power Grid Faults

We set all power lines to an open-circuit state, paralyzing the power grid in the TNCS. In this scenario, the impact of uncertainties such as failed unit start-up and the occurrence

of new faults caused by the restoration process is analyzed, and the algorithm proposed in this study is compared with other algorithms to evaluate the performance and efficacy.

Set the batch size to 128 and the total number of episodes to 5000, while the settings of the remaining experimental parameters are identical to those in the small-scale fault scenario. The simulation results are shown below.

We can observe that curve 3 gradually converged to a stable value in Figure 7, indicating that the actor network was well-trained and able to make accurate and effective decisions in large-scale power grid faults. Furthermore, the three curves showed different average rewards and fluctuations. Curve 3 exhibited the lowest average reward, and curve 3 and curve 1 showed significantly higher fluctuation than curve 2. This is due to the fact that both generator start-up failures and new faults occurring during the restoration process result in a reward value of -1. Consequently, curve 3 exhibited lower values than the other two curves. The presence of new faults increased the uncertain change of S and the difficulty of convergence, leading to higher fluctuations in curve 3 and curve 1.
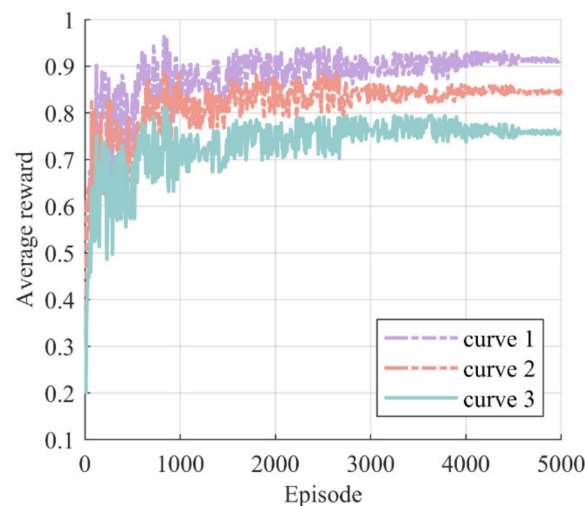

Figure 7. The impact of uncertainty factors on fault recovery. Curve 1 represents the scenario without considering the possibility of unit start-up failures during the restoration process. Curve 2 represents the scenario without considering the possibility of fault recovery causing new faults. Curve 3 represents the scenario that considers both unit start-up failures and the occurrence of new faults during the restoration process.

To validate the impact of uncertainties on the restoration sequence, we take the restoration sequence of power line 42 as an example to observe the change of its restoration sequence, and the results are shown in Figure 8. We can observe that in the case corresponding to curve 3, the restoration sequence of power line 42 in 10 experiments was approximately the 17th step, which differs from the cases corresponding to curves 1 and 2. This indicates that uncertainties during the restoration process can alter the restoration sequence and should not be ignored in the practical restoration process.

At the end of this section, we also compare the proposed algorithm with other intelligent algorithms such as particle swarm optimization (PSO), genetic algorithm (GA), and other deep reinforcement learning algorithms to verify the superiority of the proposed algorithm in solving the fault recovery problem of the TNCS [34,35]. The results are shown in Table 3. We can observe that the proposed algorithm outperformed other algorithms by 0.3% to 20.96% in terms of restoration benefit, and in terms of convergence time, there existed a reduction of 2.82% to 14.39% compared to other algorithms except for the PSO algorithm, with only a marginal increase of 1.2% compared to the PSO algorithm. The aforementioned comparative results demonstrate that the proposed algorithm has distinct advantages in terms of both restoration benefit and convergence time, indicating its superiority in an overall assessment. Moreover, compared to the TD3 algorithm, although the

restoration benefit is very close, the proposed algorithm exhibited an 8.93% reduction in time. This is due to improvements made to the TD3 algorithm, which reduces the occurrence of new faults during the restoration process and effectively reduces the uncertain change of S, thereby reducing convergence difficulties during training.
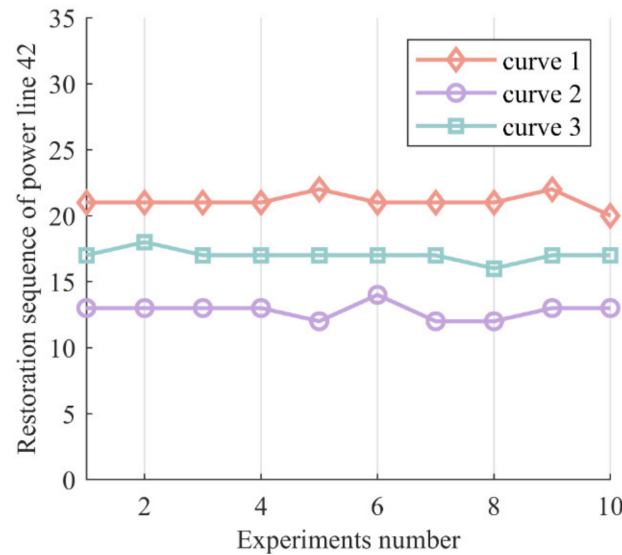
Figure 8. Impact of uncertainties on restoration sequence. The curves labeled as curve 1, curve 2, and curve 3 in this figure correspond to curve 2, curve 1, and curve 3 in Figure 7, respectively. The restoration sequence of power line 42 refers to the step at which this line is recovered.

Table 3. Comparison of different algorithms.

Algorithm	Performance Metrics	
	Restoration Benefit (MW)	**Convergence Time (s)**
Proposed algorithm	12.3968	14,763
GA	10.4519	15,192
PSO	9.7987	14,580
DQN	10.4832	17,244
AC	10.8751	16,848
DDPG	11.7285	16,416
TD3	12.3812	16,211

*3.4. Communication Faults*

In this section, we add the simulation of communication faults on the basis of large-scale power grid faults to validate the resilience of the algorithm proposed in this study.

Specifically, we assume that DoS attackers send a significant volume of disguised packets to information nodes 31, 32, 33, and 34, resulting in their infection. As a result, the communication channels within the coupling layer connected to the information nodes and the communication links within the information network of the control layer become blocked. When communication delays are detected, the technical staff at the control center suspend the restoration of power grid faults and initiate the deployment of firewalls and intrusion detection systems to restore the communication faults. Equation (25) demonstrates that the impact of communication delays on restoration benefit depends on its occurrence timing and restoration speed. To facilitate the statistical analysis, the occurrence time of communication delays is divided into six stages: stage 1: 1 to 6 steps; stage 2: 7 to 12 steps; stage 3: 13 to 18 steps; stage 4: 19 to 24 steps; stage 5: 25 to 30 steps; and stage 6: more than 30 steps. We set $H = 6$, with $h_t = 1$ corresponding to the first stage, $h_t = 2$ corresponding to the second stage, etc. There are five levels of restoration speed: level 1: $e_c = 1$; level 2: $e_c = 2$; level 3: $e_c = 3$; level 4: $e_c = 4$; and level 5: $e_c = 5$. Among them, $e_c = 1$ indicates

that the restoration of communication delay requires one step, $e_c = 2$ requires two steps, etc. We set $E = 5$ and $\alpha = 0.1$. The remaining experimental parameters are maintained in accordance with the previous section. The simulation results are shown as follows.

As shown in Figure 9, the restoration benefit of the communication delay curves was less than that of the normal communication curve, which was due to the communication delay causing a reduction in the Γ value of certain power lines, illustrated in Figure 10.
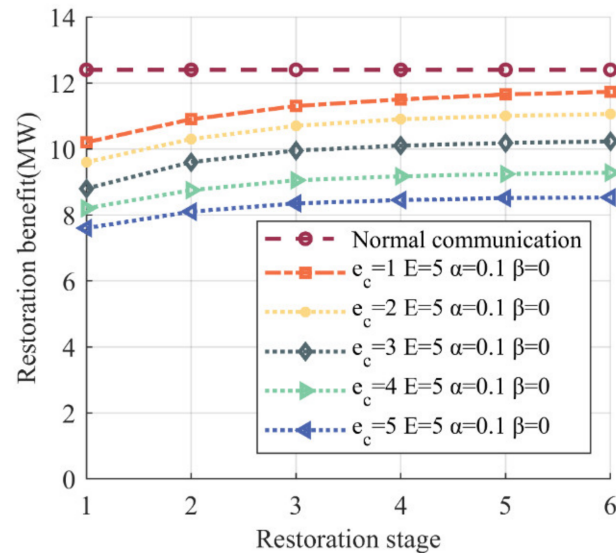

Figure 9. Impact of communication delays on restoration benefits. There is one curve representing normal communication, the value of which remains constant at 12,3968 MW, and five curves representing communication delay.

Figure 9 also illustrates two additional aspects: first is the impact of the occurrence time of DoS attacks on restoration, and second is the impact of restoration speed of communication delay on restoration.

For the first aspect, we can observe that when the restoration speed remained constant, the later the occurrence of communication delay caused by DoS attacks, and less of the restoration benefit was reduced. This is due to the fact that a delayed occurrence of communication delay results in fewer affected power lines, resulting in less reduction in the $\sum_{m}^{e_p} \Gamma$ value of the objective function (as explained by the mechanism for Γ value reduction), while the $\sum_{m}^{e_p} LS_{rm} + P_{rm}^S$ value remains unchanged. Consequently, the communication delay curve exhibits an increasing trend as the restoration process progresses.

The timing of DoS attacks is characterized by uncertainty, while the reduction in restoration benefit caused by DoS attacks can be mitigated by adjusting the value of β. According to Table 4, when the communication delay occurs in stage 1, stage 2, and stage 3, setting β to its maximum value of 1 minimizes the reduction in restoration benefit by 9.23%, 4.22%, and 0.68%, respectively, compared to the normal communication benefit of 12.3968 MW. However, in stage 4, setting β to 1 leads to a restoration benefit of 12.510 MW, which is higher than the normal benefit and unrealistic. Therefore, β cannot be set to 1, and $\beta = 0.75$ results in a restoration benefit of 12,267 MW, which is marginally less than 12,3968 MW. Thus, the valid range for β in stage 4 is $0.75 \leq \beta < 1$. Similarly, in stage 5 and stage 6, the valid range for β is $0.75 \leq \beta < 1$ and $0.5 \leq \beta < 0.75$, respectively.
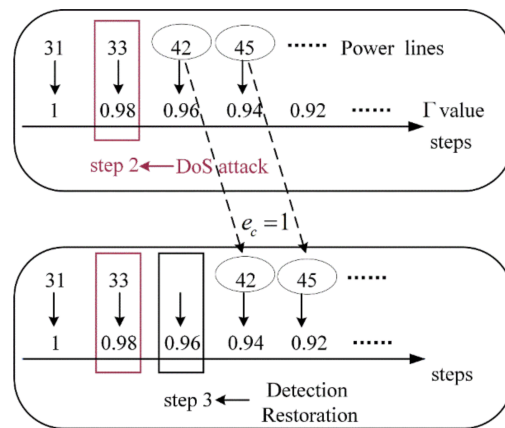
Figure 10. Mechanism of the Γ value reduction. We assume that a DoS attack is launched against the information node at step 2. It can be observed that the DoS attack has no impact on the restoration of steps 2 and 1. As the restoration process reaches step 3, the control center performs a status check and detects the communication delay, initiating its restoration. Power line 42 cannot be restored in step 3, and its recovery order is changed from step 3 to step 4. The reason it is changed to step 4, rather than step 5, step 6, etc., is because $e_c = 1$. Similarly, the restoration step for power line 45 transitions from step 4 to step 5, etc. At step 4, regardless of whether power lines 45 or 42 are restored, the Γ value corresponding to step 4 remains unchanged. When power line 42 is restored at step 4, the Γ value reduces from 0.96 (corresponding to step 3) to 0.94. The above analysis explains the reduction of the Γ value.

Table 4. The impact of β on restoration benefit.

h_t	Restoration Benefit (MW) (e_c=1)				
	$\beta = 1$	$\beta = 0.75$	$\beta = 0.5$	$\beta = 0.25$	$\beta = 0$
1	11.252	10.955	10.705	10.457	10.207
2	11.873	11.590	11.401	11.159	10.915
3	12.312	12.065	11.823	11.577	11.321
4	12.510	12.267	12.030	11.780	11.531
5	12.627	12.371	12.120	11.878	11.652
6	12.800	12.550	12.231	11.990	11.730

For the second aspect, it can be observed that the restoration benefit of the communication delay curve corresponding to $e_c = 1$ was greater than that of the curves corresponding to other restoration speeds in all restoration stages. This is due to restoring communication delay leading to obtaining the reward I_c. If the restoration speed is fast, such as completing the restoration in a single step, one I_c is obtained. Alternatively, multiple I_c can be obtained if the restoration speed is slow and requires multiple stages to complete. However, I_c and $I_b + I_m$ are not equivalent, and a difference exists between them. Obtaining multiple I_c can lead to overestimation or underestimation of $Q(S, A)$, which may result in policy bias and affect restoration decision-making. To solve this issue, we can adjust the values of β and α to reduce the difference between I_c and $I_b + I_m$. It is important to note that we cannot directly control the restoration speed due to its uncertainty and dependence on the defensive deployment level of the information network. When the occurrence time of DoS attacks is fixed, we can adjust the value of α to mitigate the policy bias caused by restoration speed. Figure 11 illustrates the impact of different values of α on policy bias when $\beta = 0.5$.
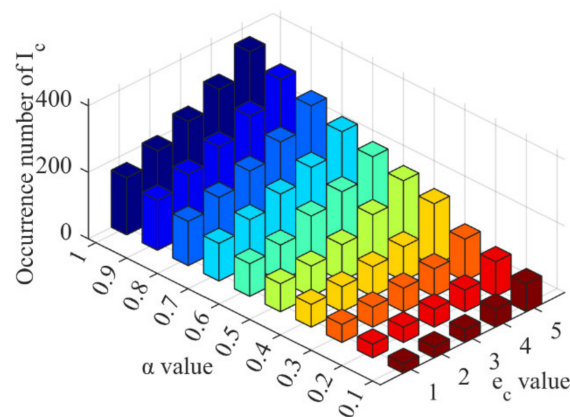
Figure 11. The impact of α on policy bias. The various colored bars represent various α values.

It can be observed from Figure 11 that as the value of α reduces, the occurrence number of I_c reduces, regardless of the restoration speed. First, this indicates that when $\beta = 0.5$, the value of I_c is greater than $I_b + I_m$, resulting in a $Q(S, A)$ overestimation. To solve this, the value of β can be appropriately reduced by considering the occurrence time of DoS attacks. Second, the minimum occurrence number of I_c reduced from 60.26% to 80.12% compared to the maximum occurrence number for the five restoration speeds, by adjusting the value of α, demonstrating that adjusting the value of α effectively resolves the policy bias issue caused by restoration speed.

The average occurrence number of I_c for the five restoration speeds under each α value in Figure 11 is calculated to represent the actual policy bias under the normal restoration speed. The results are shown in Table 5, where the difference ratio indicates the percentage reduction in the average occurrence number of I_c for the next α value compared to the current α value. For instance, when the value of α changed from 1 to 0.9, the average occurrence number of I_c reduced by 11.29%, resulting in a difference ratio of 11.29%. We can observe that as the value of α reduced from 1, the difference ratio increased continuously. When the value of α changed from 0.4 to 0.3, the difference ratio reached its maximum value of 24.38%, and as the value of α continued to reduce, the difference ratio also reduced. This indicates that under the condition of $\beta = 0.5$, the optimal range of α value for minimizing the difference between I_c and $I_b + I_m$ was around 0.3, leading to a more effective improvement in solving the problem of policy bias.

Table 5. The impact of α on policy bias.

α Value	I_c Average Occurrence Number	Difference Ratio
1	258.6	0
0.9	229.4	11.29%
0.8	202.6	11.68%
0.7	177.4	12.44%
0.6	153.2	13.64%
0.5	129.2	15.67%
0.4	104.2	19.35%
0.3	78.8	24.38%
0.2	65.0	17.51%
0.1	55.5	14.62%

Based on the aforementioned analysis, it is evident that for uncertainties that cannot be directly controlled, such as the occurrence time of DoS attack and the restoration speed of communication delays, we can adjust the value of β to mitigate the extent of restoration loss under various occurrence times of a DoS attack. Additionally, we can improve the issue of policy bias caused by restoration speed by adjusting the value of α. The experimental

results demonstrate that the proposed algorithm exhibits strong adaptability and resilience in the presence of communication delays caused by a DoS attack.

4. Conclusions

In addressing the issue of fault recovery in a TNCS, we propose a recovery method based on an improved TD3 algorithm. In the design of this method, we have particularly considered the uncertainties, system security issues, and communication faults. Experimental results indicate that the proposed method can obtain optimal recovery decisions and accomplish the maximum restoration benefit when faults occur in the power grid of the TNCS. Furthermore, the improvement to the TD3 algorithm effectively reduces the occurrence of energy supply–demand imbalance, and the uncertainties can impact the restoration sequence so that they should not be ignored during the restoration process. The results also show that communication delays caused by DoS attacks can reduce the restoration benefit and lead to policy bias. By adjusting the values of β and α, the proposed algorithm can effectively mitigate the extent of restoration loss under various occurrence times of a DoS attack and improve the issue of policy bias caused by the restoration speed, thereby demonstrating strong resilience.

From a broader application perspective, the method we propose offers a reference theoretical model for the integrated development of energy, transportation, and information. In practical applications, model parameters can be adjusted based on specific circumstances, aiding decision-makers in more scientifically informed choices. Nevertheless, the proposed approach faces challenges. For instance, the intricate operating conditions of transportation networks might necessitate targeted studies for specific situations. Moreover, when confronted with a variety of sophisticated network attacks, questions arise regarding the model's resilience and potential necessary adjustments. Therefore, regarding future research directions, several avenues can be explored, such as studying the impact of transportation network faults on the restoration process, designing recovery methods when confronted with various types of network attacks, and employing other intelligent algorithms to solve the fault recovery problem in TNCSs.

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

References

1. Yang, T.; Guo, Q.; Sheng, Y. Urban integrated electric-traffic network collaboration from perspective of system coordination. *Autom. Electr. Power Syst.* **2020**, *44*, 1–9.
2. He, Z.; Xiang, Y.; Liao, K. Demand, form and key technologies of integrated development of energy-transport-information networks. *Autom. Electr. Power Syst.* **2021**, *45*, 73–86.
3. Hu, Q.; Ding, H.; Chen, X. Analysis on rotating power outage in California, USA in 2020 and its enlightenment to power grid of China. *Autom. Electr. Power Syst.* **2020**, *44*, 11–18.
4. Sun, H.; Xu, T.; Guo, Q. Analysis on blackout in Great Britain power grid on August 9th, 2019 and its enlightenment to power grid in China. *Proc. CSEE* **2019**, *39*, 6183–6192.
5. Zhang, Y.; Xie, G.; Zhang, Q. Analysis of 2.15 power outage in Texas and its implications for the power sector of China. *Electr. Power* **2021**, *54*, 192–198.
6. Yan, D.; Wen, J.; Du, Z. Analysis of Texas blackout in 2021 and its enlightenment to power system planning management. *Power Syst. Prot. Control.* **2021**, *49*, 121–128.
7. Hu, Y.; Xue, S.; Zhang, H. Cause analysis and enlightenment of global blackouts in the past 30 years. *Electr. Power* **2021**, *54*, 204–210.

8. Wu, J.; Chen, Z.; Zhang, Y. Sequential recovery of complex networks suffering from cascading failure blackouts. *IEEE Trans. Netw. Sci. Eng.* **2020**, *7*, 2997–3007. [CrossRef]
9. Han, O.; Chen, Z.; Ding, T. Power system black-start restoration model considering wind power uncertainties. *Power Syst. Technol.* **2022**, *147*, 108855.
10. Pang, K.; Wang, Y.; Wen, F. Cyberphysical collaborative restoration strategy for power transmission system with communication failures. *Autom. Electr. Power Syst.* **2021**, *45*, 58–67.
11. Li, Z.; Xue, Y.; Wang, H. A dynamic partitioning method for power system parallel restoration considering restoration-related uncertainties. *Energy Rep.* **2020**, *6*, 352–361. [CrossRef]
12. Arif, A.; Wang, Z.; Wang, J. Power distribution system outage management with co-optimization of repairs, reconfiguration, and DG dispatch. *IEEE Trans. Smart Grid* **2017**, *9*, 4109–4118. [CrossRef]
13. Liu, T.; Zhu, Y.; Sun, R. Resilience-enhanced-strategy for cyber-physical power system under extreme natural disasters. *Autom. Electr. Power Syst.* **2021**, *45*, 40–48.
14. Sun, L.; Liu, W.; Chung, C. Improving the restorability of bulk power systems with the implementation of a WF-BESS system. *IEEE Trans. Power Syst.* **2018**, *34*, 2366–2377. [CrossRef]
15. Yang, Z.; Sun, L.; Ding, M. Optimization strategy for start-up sequence of generation units considering critical restoration paths. *Electr. Power Constr.* **2019**, *40*, 28–35.
16. Xiang, Y.; Ding, Z.; Zhang, Y. Power system reliability evaluation considering load redistribution attacks. *IEEE Trans. Smart Grid* **2016**, *8*, 889–901. [CrossRef]
17. Che, L.; Liu, X.; Shuai, Z. Cyber cascades screening considering the impacts of false data injection attacks. *IEEE Trans. Power Syst.* **2018**, *33*, 6545–6556. [CrossRef]
18. Che, L.; Liu, X.; Li, Z. False data injection attacks induced sequential outages in power systems. *IEEE Trans. Power Syst.* **2018**, *34*, 1513–1523. [CrossRef]
19. Wang, Q.; Pipattanasomporn, M.; Kuzlu, M. Framework for vulnerability assessment of communication systems for electric power grids. *J. IET Gener. Transm. Distrib.* **2016**, *10*, 477–486. [CrossRef]
20. Huseinović, A.; Mrdović, S.; Bicakci, K. A survey of denial-of-service attacks and solutions in the smart grid. *IEEE Access* **2020**, *8*, 177447–177470. [CrossRef]
21. Tan, S.; Sun, J.; Wan, L. A DoS attack intensity-aware adaptive critic design of frequency regulation for EV-integrated power grids. *Int. J. Electr. Power Energy Syst.* **2023**, *145*, 108656. [CrossRef]
22. Li, X.; Jiang, C.; Du, D. A Novel State Estimation Method for Smart Grid Under Consecutive Denial of Service Attacks. *IEEE Syst.* **2022**, *17*, 513–524. [CrossRef]
23. Zhang, B.; Dou, C.; Yue, D. Attack-defense evolutionary game strategy for uploading channel in consensus-based secondary control of islanded microgrid considering DoS attack. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2021**, *69*, 821–834. [CrossRef]
24. Gao, Q.; Du, Z.; Gi, Y. Resilient load frequency control for multi-area interconnected power system under denial-of-service attacks. *Electr. Power Constr.* **2023**, *44*, 54–62.
25. An, X.; Sun, H.; Zhang, X. Analysis and lessons of Texas power outage event on February 15, 2021. *Proc. CSEE* **2021**, *41*, 3407–3415.
26. Li, M.; Sun, L.; Ma, Y. An optimization strategy for generator start-up sequence after blackouts considering the cyber system fault. *Electr. Power* **2022**, *55*, 146–155.
27. Fujimoto, S.; Hoof, H.; Meger, D. Addressing function approximation error in actor-critic methods. In Proceedings of the International Conference on Machine Learning. PMLR, Stockholm, Sweden, 9 February 2018; pp. 1587–1596.
28. Xie, S.; Girshick, R.; Dollár, P. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
29. Wang, Q.; Wu, B.; Zhu, P. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
30. Mnih, V.; Kavukcuoglu, K.; Silver, D. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]
31. Sutton, R.S.; McAllester, D.; Singh, S. Policy gradient methods for reinforcement learning with function approximation. *Adv. Neural. Inf. Process. Syst.* **1999**, *12*, 1057–1063.
32. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
33. Liu, Y.; Gu, X. Skeleton-network reconfiguration based on topological characteristics of scale-free networks and discrete particle swarm optimization. *IEEE Trans. Power Syst.* **2007**, *22*, 1267–1274. [CrossRef]
34. Del Valle, Y.; Venayagamoorthy, G.K.; Mohagheghi, S. Particle swarm optimization: Basic concepts, variants and applications in power systems. *IEEE Trans. Evol. Comput.* **2008**, *12*, 171–195. [CrossRef]
35. Golpira, H.; Bevrani, H. Application of GA optimization for automatic generation control design in an interconnected power system. *Energy Conv. Manag.* **2011**, *52*, 2247–2255. [CrossRef]