Resilience Assessment for Power Systems Under Sequential Attacks Using Double DQN with Improved Prioritized Experience Replay

Lingkang Zeng, Wei Yao, *Senior Member, IEEE*, Hang Shuai, *Member, IEEE*, Yue Zhou, *Member, IEEE*, Xiaomeng Ai, *Member, IEEE* and Jinyu Wen, *Member, IEEE*

***Abstract*—The information and communication technology enhances the performance and efficiency of cyber-physical power systems (CPPSs). However, it makes the topology of CPPSs more exposed to malicious cyber attacks in the meantime. This paper proposes a double deep-Q-network (DDQN) based resilience assessment method for power systems under sequential attacks. The DDQN agent is devoted to identifying the least sequential attacks to the ultimate collapse of the power system under different operating conditions. A cascading failure simulator considering the characteristics of generators is developed to avoid a relatively optimistic assessment result. In addition, a novel resilience index is proposed to reflect the capability of the power system to deliver power under sequential attacks. Then, an improved prioritized experience replay technique is developed to accelerate the convergence rate of the training process for DDQN agent. Simulation results on the IEEE 39-bus, 118-bus and 300-bus power systems demonstrate the effectiveness of the proposed DDQN-based resilience assessment method.**

***Index Terms*—Resilience assessment, cyber-physical power system, cascading failure simulation, double deep-Q-network, prioritized experience replay.**

I. Introduction

ADVANCED information and communication technology has been integrated into power systems for the enhancement in performance and efficiency, which turns electrical grids into cyber-physical power systems (CPPSs) [1]–[3]. However, the complex interconnectivity between different devices and elements makes CPPSs more exposed to malicious cyber and physical attacks [4], [5]. Specifically, the topology attacks could cause the power line outages, which would lead to the widespread power flow transfer and further cascading failures [6]. It can create severe damage in CPPSs, for example, the 2015 Ukraine blackout caused by false data injection attacks [7]. Hence, it is necessary to conduct the resilience assessment in advance [8], in order to provide guiding suggestions for the investment in resilience-enhancement countermeasures [9], [10].

In spite of the not-yet-standardized definitions of resilience of CPPSs [11], there is a relatively widely recognized definition: the ability of CPPSs to tolerate cyber-based and power-based disturbances or recover from disturbances by operation technology [12]. There have been plenty of researches on the resilience assessment of power systems under power-based attacks or extreme weather events [13], [14]. However, compared with the power-based faults, the cyber-based attacks on topology require less attack resources and can be conducted more flexibly. It can also lead to severe blackouts [6], [15], [16], which is launched by tampering the status data [17], [18] or malicious line-switching operation [19], [20]. Usually, there are two main kinds of topology attacks: synchronous attack [21]–[23] and sequential attack [24]–[27]. The sequential attack is proposed in [24] firstly and proved that it may result in severer blackouts than the synchronous attack. Besides, it requires less coordination on attack resources, which means the higher flexibility in the choice of attack lines [27]. Hence, this paper will focus on the resilience assessment of CPPSs under the sequential cyber attacks on the power topologies.

Generally, the resilience assessment of power systems identifies the least sequential attacks for the system failure. It consists of two main parts: the cascading failure simulator (CFS) and the design of contingency/attack. CFS is proposed to simulate the cascading failure and corresponding recovery operation strategies of the power systems under attack/disturbance [28]. Basically, they are designed using the dc power flow (DC-PF) model [29]–[32] or the computationally efficient AC model [33]. Some stochastic cascading failure models [34], [35] are further proposed to simulate the cascading failures considering the high uncertainty of load and renewable generations. However, the characteristics of generators are sometimes not sufficiently considered for the generator rescheduling during the cascading failure process in the above researches. For example, Ref. [29]–[31] ignored the lower power limits of generators. Ref. [33]–[35] ignored the governor droop coefficients, which reflected the primary frequency regulation. Ref. [32] considered the characteristics of generators, but ignored the power loss when deciding the load shedding. These insufficient considerations might lead to a relatively optimistic result of the resilience assessment under sequential attacks. The post-contingency evolution simulator proposed in [36] made up for these deficiencies. Whereas, due

Manuscript received December 9, 2021; revised February 5, 2022; accepted April 26, 2022. This work was supported by National Natural Science Foundation of China under Grant U1866602. Paper no. ISJ-RE-21-13442 *(Corresponding author: Wei Yao.)*

L. Zeng, W. Yao, X. Ai and J. Wen are with State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronics Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China.(Email: lingkang.zeng@foxmail.com; w.yao@hust.edu.cn; xiaomengai@hust.edu.cn; jinyu.wen@hust.edu.cn)

H. Shuai is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA.(Email: hangshuai523@gmail.com)

Y. Zhou is with the School of Engineering, Cardiff University, Cardiff CF24 3AA, Wales, UK.(Email: ZhouY68@cardiff.ac.uk)

to the ac power flow and various modelled control measures, it is too computationally intensive for the resilience assessment under sequential attacks. Moreover, there are few resilience indexes to describe the functionality of the power system to deliver power under sequential attacks [11].

On the other hand, the contingency/attack for resilience assessment is usually designed by operation experience. The resilience of critical electrical power infrastructure is analyzed under extreme weather events in [37], or evaluated by Monte Carlo simulation in [38]. However, the rarely occurred extreme weather events and blackouts caused by cyber-attacks in practice may lead to the scarcity of historical data. Besides, numerous simulations would bring a considerable computational burden in modern complex power systems.

With the rapid development of reinforcement-learning (RL) theory [39], [40], some RL algorithms are introduced to design the contingency/attack for the resilience assessment of power systems. The RL-based agent can be trained with the experience data through the interaction with CFS, rather than the historical data of extreme weather events. Ref. [41] proposed a table-based Q-learning method to identify the critical sequential topology attacks. However, the branch operating state is discrete, which means the Q-table needs to be retrained for another new operating condition. Thanks to the development of deep learning in recent years, the deep neural network can be used to evaluate actions with respect to continuous states [42]. A deep-Q-network (DQN) based cyber-physical coordinated attack strategy is proposed in [43] to attack the critical line with the minimal cyber resources. Nevertheless, the DQN agent in [43] still cannot deal with different operating conditions due to the same state design as [41]. Moreover, the threshold of practical grids collapse might be higher than that in [41], [43]. It means that the Markov Decision Process (MDP) lasts relatively longer, and the action space would be huge. As a result, a much larger Q-table is required to save the Q value of different state-action pairs. Besides, numerous transitions with low rewards would be generated by the exploration. It would make the DQN agents trained with the conventional experience replay technique spend a lot of time converging to the approximate optimal, especially when the computing and training resources are limited.

Considering the aforementioned problems, this paper proposes a double deep-Q-network (DDQN) based method for resilience assessment of power systems under sequential attacks using improved prioritized experience replay (PER). The proposed DDQN based agent is supposed to identify the least sequential attacks resulting in the system collapse under different operating conditions. Numerous simulation results on the IEEE 39-bus, 118-bus and 300-bus power systems verify the effectiveness of the proposed DDQN based agent for resilience assessment under sequential attacks.

The main contributions of this work are summarized here:

- A DDQN based agent using an improved PER technique is proposed for the resilience assessment of power systems under sequential attacks among different operating conditions.
- A DC-PF based CFS is proposed to simulate the complex failure and corrective manipulations of power systems under sequential attacks, which can consider the characteristics of generators. In addition, a novel resilience index is designed to reflect the capability of CPPSs to deliver power under sequential attacks.
- The improved PER technique is proposed to accelerate the convergence of the training process for DDQN based agents. It scores the priority of experience in replay buffer with the combination of temporal difference (TD) error and the average episode reward.

The rest of this paper is organized as follows. The CFS considering generator characteristics is briefly introduced in Section II. Section III presents the MDP formulation of the DDQN based agent for resilience assessment followed by the introduction of improved PER technique. Section IV discusses the simulation results on three IEEE benchmarks and Section V concludes the paper.

II. Cascading Failure in Power Systems under Sequential Attacks

A. Cascading Failure Simulator Under Sequential Attacks

Malicious line-switching attacks on critical lines can turn their status from in-service to out-of-service. It can trigger large-scale power flow transfer and lead to severe cascading failure. When the number of out-of-service lines reaches a certain threshold, it can be regarded as a collapse of the power system [41]. After that, the line maintenance and power recovery will take a long time. Referring to the architecture of the MATCASC toolbox in MATLAB [44], the cascading failure simulator based on the DC-PF model is shown in Fig. 1. It is expected to make the power system collapse through the least topology attacks during the resilience assessment. To this end, the malicious sequential topology attacks and the cascading failures resulting from the attacks are conducted in turn. Note that the branch short time emergency ratings of transmission lines are set as the thermal capacity and their augmented values. The detailed information refers to [6]. The procedures of the cascading failure simulation based power system resilience assessment under sequential attacks are as follows:

1) Load the initial operating condition of the power system, including the bus data, line data, and generator data.
2) Calculate the power flow on lines with the DC-PF model.
3) Attack one of the operating lines and update the topology of the power system.
4) Detect whether the whole power system turns into several sub-grids after the attacked line is tripped out. All the lines in the sub-grid without generators turn into out-of-service state.
5) Re-dispatch the active generations and loads to ensure the active power balance in each sub-grid.
6) Update the power flow with the re-dispatched data using the DC-PF model. If there exist overloaded lines, trip the overloaded lines with minimal capacity and turn to 4).
7) Check whether the total line-outages N^{total} are less than the threshold N_{c}. If yes, turn to step 3) and attack another

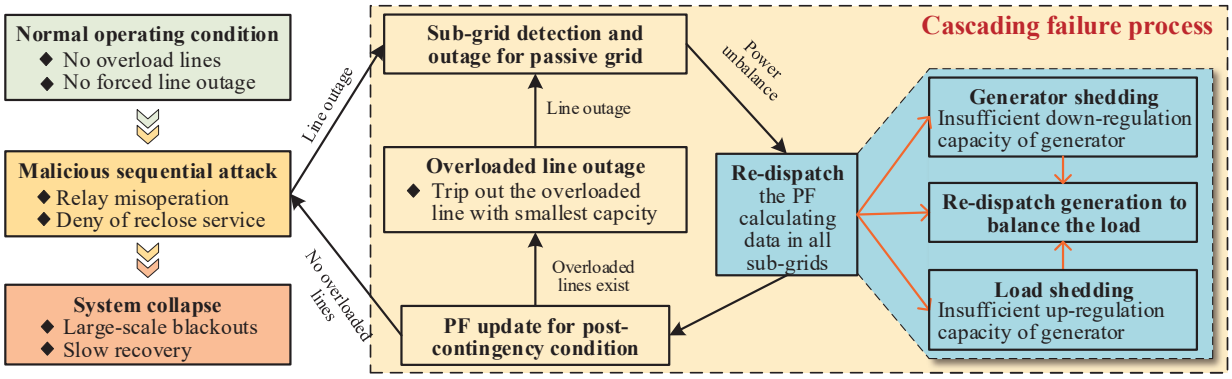


Fig. 1. System collapse caused by sequential attacks and cascading failure

operating line. Otherwise, output the attack sequence. The number of attacks will reflect the resilience of the power system under this operating condition.

B. Re-dispatch Strategy During Cascading Failure

Different from the *Cascading Failure Simulation* module in [44], the re-dispatch strategy in this paper additionally considers the governor droop coefficient and the lower power limits of generators. The details of the re-dispatch strategy are given as follows:

1) ***Calculate unbalanced power***: Summarise the total load P_L, active power of each generator P_g^i and the corresponding upper limit $P_{\max}^i$, and lower limit $P_{\min}^i$ in the concerned sub-grid. Calculate the unbalanced power ΔP as follow:

$$\Delta P = P_L - \Sigma P_g^i \tag{1}$$

2) ***Load shedding***: If $\Delta P > \Sigma(P_{\max}^i - P_g^i)$, the power shortfall will still exist after increasing all the generations to their upper limits. In this situation, cut off the load in amount of ΔP_L:

$$\Delta P_L = 0.1 P_L \lceil \frac{\Delta P - \Sigma(P_{\max}^i - P_g^i)}{0.1 P_L} \rceil \tag{2}$$

It means that the load would be cut off at internals of 10% of the current load. Although the CFS in this paper is based on the DC-PF model, the $\lceil \bullet \rceil$ function considers the power loss on the transmission lines in practical grids. Then the unbalanced power goes to $\Delta P = P_L - \Delta P_L - \Sigma P_g^i$.

3) ***Generator tripping***: If $|\Delta P| > \Sigma(P_g^i - P_{\min}^i)$, the power surplus will still exist after decreasing all the generations to their lower limits. In this situation, trip off the generator in ascending order of $(P_g^i - P_{\min}^i)$ until $|\Delta P| \le \Sigma(P_g^i - P_{\min}^i)$. Then re-calculate the unbalanced power ΔP. This is because the heat generated by the friction of valve and steam cannot be taken away in time at a low operating point. It could cause the damage to the blade of steam turbine of generator when the generator operates under the lower power limit. It is also uneconomical for generators to operate at a fairly low power point with the consideration of generating cost, like auxiliary power consumption.

4) ***Power generation adjustment***: Adjust the power of each operating generator according to its droop coefficient R_i. During the primary frequency regulation for the unbalanced power, the governor droop coefficient approximately determines how much the generator responds to the frequency deviation [36]. Note that the frequency dynamics is ignored here. The power adjustment is formulated as follows:

$$\begin{cases} \Delta P_g^i = \dfrac{1}{R_i} \Delta P / \Sigma \dfrac{1}{R_i} \\ P_{\min}^i < P_g^i < P_{\max}^i \end{cases} \tag{3}$$

If P_g^i reaches its own upper or lower limit, the rest adjustment amount will be apportioned to other generators.

C. Resilience Index Based on the Number of Attacks

Since the resilience of the power system in this paper reflects the capacity to tolerate malicious attacks and continue to deliver affordable power, the resilience index is defined based on the number of attacks as follows:

$$\begin{cases} RI_k = \sum_{k=1}^{k_e} F(k)^{1/k} \\ F(k) = (N - N_k^{\text{total}})/N \\ 1 \le k \le k_e, k \in \mathbb{N}^+, k_e \in \mathbb{N}^+ \end{cases} \tag{4}$$

where RI_k is the resilience index of power system after k attacks. N is the number of operating lines in the initial operating condition, while N_k^{total} is the total out-of-service lines caused by k attacks. $F(k)$ is the ratio of the number of the remaining operating lines to the number of the initial operating lines, which reflects the damage to the structure of the power system after k attacks. Note that $F(k)$ is positive and in the range of $[0,1)$, while k is a positive integer. k_e is the number of attacks when the system collapses ($N_k^{\text{total}} \ge N_c$).

Since RI_k is the cumulative value of $F(k)$ to the power of $1/k$, it increases with the attack times k until the power system collapses. It is reasonable that the more attacks the power system can tolerate, the more resilient the system is. In addition, the power operator is no greater than 1, which is a kind of amplification of $F(k)$ in the integration. The amplification increases with the attack times k. It means that the capability to tolerate multiple attacks are more precious and counts more in the resilience assessment of power system.

III. Deep Reinforcement Learning Based Resilience Assessment

The DDQN based resilience assessment method is introduced in the following section. Firstly, the resilience assessment of the power system under sequential attacks will be formulated as an MDP problem. Then, the DDQN agent using the improved PER technique is developed to make decisions of sequential attacks. The end of this section gives the training procedures of the DDQN based agent for resilience assessment.

A. Markov Decision Process Formulation of Resilience Assessment under Sequential Attacks

Fig.2 shows the MDP of the resilience assessment for the power system which reflects the interaction between the agent and the proposed cascading failure simulator. The key elements of MDP are state s_k, action a_k, reward r_k, terminal T_{end} and the state transformation function. In detail, the agent firstly decides the attacking line a_k when it received the current power flow state s_k from the concerned power system. Then the DC-PF based cascading failure simulation is conducted with the state s_k and action a_k until there is no overloaded line. That represents the state transformation completes and the next power flow state s_{k+1} is obtained. The corresponding reward is made according to the reward function which is designed in prior. The MDP continues as the current state s_k is replaced by s_{k+1}, until the terminal state of the MDP arrives and T_{end} is true.

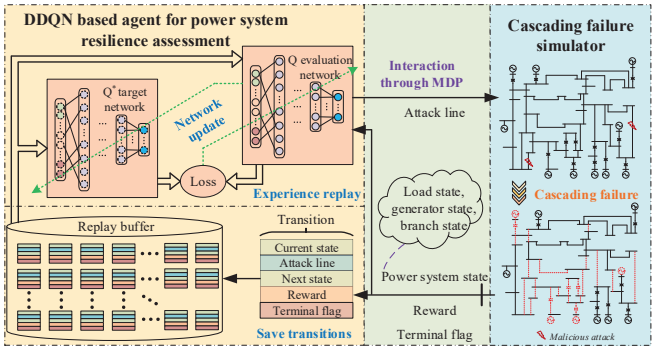


Fig. 2. The framework of the DRL based resilience assessment method

1) ***State***: The state s_k refers to the power flow state of power system at decision step k, which can be described as:

$$s_t = \left[P_{\text{L}}^1(k),...,P_{\text{L}}^x(k),P_{\text{g}}^1(k),...,P_{\text{g}}^m(k),P_{\text{b}}^1(k),...,P_{\text{b}}^y(k)\right] \quad (5)$$

where $P_{\text{L}}^x(k)$ represents the active power load of bus x at step k. $P_{\text{g}}^m(k)$ represents the active power generation of generator m at step k. $P_{\text{b}}^y(k)$ represents the active power transmitted on the branch y at step k. x, m and y are the number of the observed buses, generators and branches, respectively.

2) ***Action***: The action a_k refers to the index of the power transmission lines that the agent decides to attack at decision step k. Accordingly, a_k can be described as:

$$\begin{cases} a_k \in \boldsymbol{A_k}, \boldsymbol{A_k} = \left\{i | i \in \boldsymbol{A_0} \cap B_k^i = 1\right\} \\ \boldsymbol{A_0} = \{1, i, ..., N\} \\ \boldsymbol{B_k} = \left[B_k^1, B_k^i, ..., B_k^N\right], B_k^i \in \{0,1\} \end{cases} \quad (6)$$

where, $\boldsymbol{A_k}$ represents the collection of available attacking lines in power system at decision step k. $\boldsymbol{A_0}$ represents the initial collection at step 0 when all the transmission lines are under operation. N is the corresponding number of lines in the initial collection which is also the dimension of the output layer of the agent. $\boldsymbol{B_k}$ refers to the operating state of all the branches listed in $\boldsymbol{A_0}$ and B_k^i is the operating state of branch i at decision step k, where 0 means out-of-service and 1 means under operation. In this way, the transmission line selected by the agent at every decision step is guaranteed under operation.

In addition, the ε-greedy strategy is also adopted for action selection, which is beneficial to make a balance between exploitation and exploration during the training process. It is formulated as follows:

$$\begin{cases} P(a_k = a_k^*) = 1 - \varepsilon_k, a_k \in \boldsymbol{A_k} \\ a_k^* = \arg\max_{a_k \in A_k} Q(s_k, a_k) \end{cases} \quad (7)$$

where a_k^* is the action with the biggest value evaluated by Q network. $P(a_k = a_k^*)$ is the probability of selecting the most valued action at decision step k, which is $1 - \varepsilon_k$. It means that there is a probability of ε_k for randomly selecting a transmission line in $\boldsymbol{A_k}$.

$$\varepsilon_k = min\left\{\varepsilon_0 - n_s \times \Delta\varepsilon, \varepsilon_f\right\} \quad (8)$$

where ε_0 and ε_f are the initial and final value of ε. n_s is the training step and $\Delta\varepsilon$ is the attenuation of the ε. Usually, $1 \geq \varepsilon_0 \geq \varepsilon_k \geq \varepsilon_f \geq 0$. Note that the ε-greedy strategy is only applied in the training stage. For the action selection in the evaluation stage, the greedy strategy is adopted instead.

3) ***Reward*** and ***Terminal***: The reward $r_k(s_k, a_k)$, referred to r_k, is designed according to the state transition and the terminal flag. The formulation of r_k is obtained as:

$$r_k = -k_1 + k_2 \times N_k^o + k_3 \times T_{\text{end}} \quad (9)$$

where k_1 and k_2 are the coefficients for the attack cost and effect, respectively. k_3 is the reward when the MDP terminates. The attack effects N_k^o refers to the number of newly increased transmission lines that are out of service after the cascading failure simulation under the attack a_k. It can be obtained as:

$$N_k^o = \sum_{i=1}^{N} B_k^i - \sum_{i=1}^{N} B_{k+1}^i \quad (10)$$

The terminal T_{end} can be formulated as:

$$T_{end} = \begin{cases} 0, N_k^{\text{total}} < N_{\text{c}} \\ 1, N_k^{\text{total}} \geq N_{\text{c}} \end{cases} \quad (11)$$

where N_k^{total} is the number of line outages, $N_k^{\text{total}} = N - \sum_{i=1}^{N} B_{k+1}^i$. N_c is the threshold of the out-of-service lines that represents the power system collapse. When the total out of service lines are less than N_{c}, T_{end} equals 0. It means that the MDP has not yet terminated and the agent will continue deciding the attack transmission line. Otherwise, T_{end} equals 1, which means the MDP has terminated and the resilience assessment is completed. The number of attacks in the sequence $[a_1, a_2, ..., a_k]$ reflects the resilience of the power system under this operating condition.

B. Double Deep-Q-Network Algorithm

The deep Q network (DQN) algorithm uses the Q-target network to not only select the action at the next state a_{k+1} but also evaluate the corresponding Q value $Q(s_{k+1}, a_{k+1})$. It could result in the over-high evaluation of the Q value at the next decision step. To this end, the double deep Q network is proposed in [42] to separate the selection of a_{k+1} and evaluation of Q_{k+1}. It can be described as:

$$Q_j^i(s_{k+1}, a_{k+1}) = Q_j^i(s_{k+1}, \underset{a_{k+1} \in \boldsymbol{A}_{k+1}}{\arg\max}\ Q_j^{*i}(s_{k+1}, a_{k+1}|\theta^{*i})|\theta^i) \tag{12}$$

where Q^i and θ^i represent the evaluation network and its parameters at i-th iteration, respectively. Q^{i*} and θ^{i*} represent the target network and its parameters at i-th iteration, respectively. a_{k+1} refers to the attacking transmission line at the next decision step. j is the number of transitions. As shown in (12), a_{k+1} is firstly determined by maximizing the output Q value of Q target network with state s_{k+1}. Then the Q value for (s_{k+1}, a_{k+1}) is evaluated by Q evaluation network. In this way, the training process of the Q evaluation network will be more stable. Note that, the ε-greedy strategy is not used during the parameters update with replayed transitions.

The purpose of training is to minimize the difference between evaluated Q value $Q_j^i(s_k, a_k|\theta_i)$ and target Q value $\widehat{Q}_j^i(s_k, a_k)$, which is represented by temporal-difference (TD) error $e_{\mathrm{TD}}(\theta^i)$:

$$e_{\mathrm{TD}_j}(\theta^i) = Q_j^i(s_k, a_k|\theta^i) - \widehat{Q}_j^i(s_k, a_k) \tag{13}$$

$$\widehat{Q}_j^i(s_k, a_k) = \left[r_j + (1 - T_{\mathrm{end}}) \times Q_j^i(s_{k+1}, a_{k+1})\right] \tag{14}$$

where $Q^i(s_{k+1}, a_{k+1})$ obtained in (12) is the evaluated Q value of the next attack. Note that, if the MDP terminates after decision step k, the target Q value $\widehat{Q}_k^i(s_k, a_k)$ equals to the reward $r(s_k, a_k)$. Then, an *Adam* optimizer is adopted to update the parameters of the Q evaluation network to minimize the mean-square TD-errors of the replayed batch of cascading failure transitions. Thus, the gradient of the network parameter is obtained as:

$$\Delta_j(\theta^i) = e_{\mathrm{TD}_j}(\theta^i) \cdot \nabla_\theta Q_j^i(s_k, a_k|\theta^i) \tag{15}$$

where $\Delta_j(\theta^i)$ is the gradient for Q evaluation network parameter θ^i brought from transition j at i-th iteration.

C. Improved Prioritized Experience Replay

The cascading failure transitions are saved in the replay buffer in form of (s_k, $\boldsymbol{B_k}$, a_k, s_{k+1}, $\boldsymbol{B}_{k+1}$, r_k, T_{end}). They will be replayed to update the parameter of Q evaluation network, which improves the exploitation of transitions and decreases the relevance of these cascading failure transitions. Thanks to the reward design in (9), the actions that cannot terminal the MDP of resilience assessment still get non-zero reward. It means that the hindsight experience replay proposed for the sparse-reward MDP problem is not needed here [45].

In general, the experience replay adopts uniform random sampling (URR), which means the sampling probability of each cascading failure transition is similar. However, those transitions with higher TD errors apparently are more unexpected to the agent and should be sampled with higher probability [46]. In [47], the transitions with higher rewards are replayed with higher probability in the training of the DRL based agent for automatic generation control (AGC) dispatch. When it comes to the resilience assessment of the power system, the transitions from the shorter attacking sequences, rather than the transitions that terminal the MDPs, are more valuable for the training of resilience assessment agents. Hence, the improved PER, which scores the priority of transition with the combination of TD-error and the average episode reward, is introduced here to accelerate the convergence rate of the training for DDQN based resilience assessment agents.

Fig. 3 shows the diagram of the improved PER, which includes stochastic prioritization replay, importance sampling weight correction, priority update, and transition replacement. Concretely, it is designed as follows:

1) ***Stochastic prioritization replay***: The stochastic prioritization replay samples a batch of transitions from the replay buffer according to their probabilities, which depend on the priorities of transitions. The priority of the cascading failure transition is formulated as:

$$\begin{cases} p_j = \max\{\eta^\alpha, (|e_{\mathrm{TD}_j}| + \overline{R_j})^\alpha\} \\ \overline{R_j} = R_{Tj}/N_{Tj} \end{cases} \tag{16}$$

where p_j is the priority of cascading failure transition j. $P_j = p_j / \sum_k p_k$ is the sampling probability of transition j, which is proportional to the priority p_j. $\sum_k p_k$ is the sum of the priorities of all transitions in replay buffer. $|e_{\mathrm{TD}_j}|$ and $\overline{R_j}$ are the absolute value of the TD error and reward of transition j, respectively. R_{Tj} and N_{Tj} are the total reward and the number of attacks of the episode that the transition j belongs to. η is the upper limit of priority, which is set to restrict the replay frequency of those transitions with extremely high priority. α is the priority coefficient representing how much the prioritization get involved in experience replay. Especially, the stochastic prioritization replay becomes uniform random replay when $\alpha = 0$.

2) ***Importance-sampling weight correction***: Since the distribution of the cascading failure transitions replayed by PER is different from that of URR, the stochastic prioritization replay inevitably introduces bias to the training process. In order to compensate for the bias, the importance-sampling (IS) weight is introduced to correct the gradient of the replayed transitions.

$$w_j = (P_j / \min\{P_j | 1 \le j \le n_{\mathrm{batch}}\})^{-\beta} \tag{17}$$

where w_j is the IS weight of cascading failure transition j. $\min\{P_j | 1 \le j \le n_{\mathrm{batch}}\}$ is the minimal priority of the replayed batch of transitions, while n_{batch} is a training hyper-parameter, the batch size. β is the coefficient of the IS weight correction , which represents how much the gradients resulting from the replayed transitions get corrected. β is set as $0 \le \beta \le 1$. Obviously, there is no IS correction for $\beta = 0$, while the IS correction is fully conducted for $\beta = 1$.

It can be seen in (17) that the higher the transition priority P_j, the lower the IS weight of the transition w_j, the more

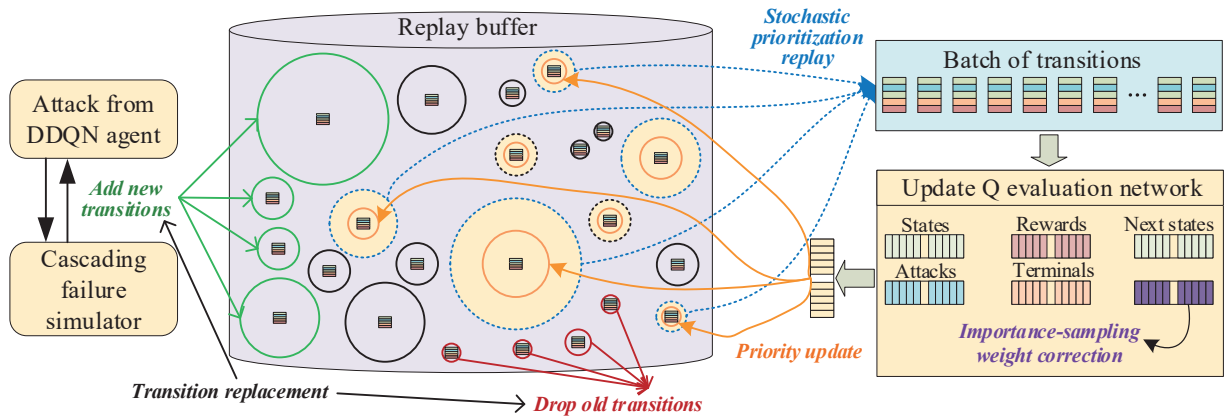


Fig. 3. The prioritized experience replay adopted in the DDQN training process (The size of circles around transitions represents the priority. The transitions in yellow are sampled as a batch to update the Q evaluation network. Then the priority of these transitions is recalculated by the latest updated networks according to (13) and (16). And the corresponding circles shrink from dashed ones to solid ones.)

the corresponding gradient magnitude decays. In this way, the gradient of the network parameter is corrected as:

$$\Delta_j(\theta^i) = w_j \cdot e_{\mathrm{TD}_j}(\theta^i) \cdot \nabla_\theta Q_j^i(s_k, a_k|\theta^i) \tag{18}$$

3) ***Priority update*** and ***transition replacement***: As shown in Fig. 3, once the parameters of the Q evaluation network is updated with the sampled cascading failure transitions, the TD errors of this batch transitions need to be re-calculated with the latest Q evaluation network. The overall decline in the priority of these transitions is foreseeable owing to the shrink of TD errors.

On the other hand, when the replay buffer is full of transitions, the old ones with the lowest priority will be replaced by the newly generated transitions. Thanks to the introduction of the average episode reward, the valuable transitions even if replayed frequently could be kept to some degree.

D. Training Process of the DDQN Based Resilience Assessment Method

The DDQN based model needs to be well-trained offline before it is used to assess the resilience of the power system under different operating conditions (OCs). The pseudo-code of the training process is given in ***Algorithm*** 1, which is conducted through the interaction with cascading failure simulator, experience replay, and network parameters update. With the performance of the agent converging gradually, the output sequential attacks will assess the resilience of the power system more accurately.

IV. Case Study

Simulations are conducted on IEEE 39-bus, 118-bus and 300-bus power systems to verify the effectiveness of the proposed DDQN based method for the resilience assessment of power systems under sequential attacks. The detailed data of these test benchmarks refer to Matpower [48].

The IEEE 39-bus power system, which consists of 10 generators, 39 buses, 21 loads, 12 transformers and 34 branches.

Algorithm 1 Training process of DDQN based agent for resilience assessment

1: Generate different initial OCs and divide them into training collection C_t and evaluation collection C_e.
2: Set total training episodes N_T, transition capacity M, training batch size B, learning rate l_r, network synchronization frequency f_s, agent evaluation frequency f_e, discount factor γ, greedy coefficient ε and other parameters.
3: Initialize the cascading failure simulator (CFS); initialize the networks of the agent; set the training step $n_s = 0$.
4: **while** $n_s \leq N_T$ **do**:
5: Update greedy coefficient ε.
6: Randomly select an initial OC in C_t;
7: s_1, $\boldsymbol{B_1} \leftarrow$ ***CFS.reset()***. ▷ *Reset the CFS and get the initial state s_1 and branch operating state $\boldsymbol{B_1}$.*
8: Set total reward $R_T = 0$, terminal flag $T_{\text{end}} = False$, attacking sequence $\boldsymbol{a} = [\ ]$.
9: **while** $T_{\text{end}} = False$ **do**:
10: $a_t \leftarrow$ ***agent.actionSelect***$(s_t, \boldsymbol{B_t}, \varepsilon)$. ▷ *Decide the attacking line according to (7).*
11: $\boldsymbol{a} \leftarrow [\boldsymbol{a}, a_t]$.
12: s_{t+1}, $\boldsymbol{B}_{t+1}$, r_t, $T_{\text{end}} \leftarrow$ ***CFS.simulate***(a_t).
13: $s_t \leftarrow s_{t+1}$, $\boldsymbol{B_t} \leftarrow \boldsymbol{B}_{t+1}$, $R_T \leftarrow R_T + r_t$.
14: **end while**
15: Output attacking sequence $\boldsymbol{a}$.
16: Calculate priorities of transitions according to (16).
17: ***transition.add()***. ▷ *Add the transition (s_t, $\boldsymbol{B_t}$, a_t, s_{t+1}, $\boldsymbol{B}_{t+1}$, r_t, T_{end}) into the replay buffer.*
18: ***transition.prioritizedReplay()***.
19: $\theta^i \leftarrow \theta^i + \Delta_j(\theta^i)$. ▷ *Update θ^i according to (18) with Adam optimizer.*
20: ***transition.update()***. ▷ *Update the priorities of the replayed transitions with the latest θ^i.*
21: $n_s \leftarrow n_s + 1$.
22: $\theta^{*i} \leftarrow \theta^i$. ▷ *Update θ^{*i} every f_s times for which θ^i updates.*
23: ***agent.eval()***. ▷ *Evaluate the agent with all OCs in C_e every f_e times for which θ^i updates.*
24: **end while**
25: Save the best and latest agent.

Note that the short time emergency ratings of those transformers connecting generators are designed according to the capacity of the connecting generators. All the loads are divided into three groups except for load 9, 12 and 31, which are less than 10 *MW* in the standard OC. The load level of each group is set to be 0.8, 0.9, 1.0 and 1.1, respectively. Then 64 different initial OCs in total are obtained, among which 54 OCs are randomly selected as training collection C_{t} and the rest ten OCs are regarded as the evaluation collection C_{e}.

A. Comparison of Cascading Failure Simulators

In order to verify the important role that the governor droop coefficient and the lower power limit of generators play in cascading failure simulation, two types of CFS are designed as follows:

1) ***CFS1***: the proposed CFS in this paper.

2) ***CFS2***: the same as CFS1 without consideration of the governor droop coefficient and the lower power limit of generators.

Table I gives the shortest attack sequences resulting in power system collapse, which are obtained by traversal method with CFS1 under all evaluation operating conditions C_{e}. Note that the threshold for the collapse of the 39-bus power system is set to 50%, which is 23 lines being out of service. The sequential attacks are conducted on CFS1 and CFS2 respectively. The resilience indexes and the corresponding numbers of lines out of service after every attack are also shown in Table I. The power system adopting CFS1 can collapse within three attacks under all C_{e}, except for *OC*.1 under which only two attacks can make it. In contrast, the power system adopting CFS2 cannot collapse after the sequential attacks under any C_{e}. The resilience indexes verify that the assessment of the proposed CFS1 is relatively more conservative. Fig. 4 shows the comparison of the lines out of service resulting from the shortest sequential attack simulated by CFS1 and CFS2.

TABLE I
THE RESULTS OF RESILIENCE ASSESSMENT CONDUCTED WITH CFS1 AND CFS2 UNDER EVALUATION OPERATING CONDITIONS C_{e}

OC.	Load level	Attacked lines	CFS1		CFS2	
			N_k^{total}	RI_3	N_k^{total}	RI_3
1	[1.0, 0.8, 0.8]	⑩-⑯	1-**24**	1.67	1-16	1.79
2	[0.9, 1.1, 0.9]	⑯-㉟-⑩	1-6-**23**	2.70	1-6-17	2.77
3	[0.9, 0.8, 0.9]	⑯-㉟-⑩	1-6-**23**	2.70	1-6-17	2.77
4	[0.8, 1.1, 0.8]	④-⑪-⑨	1-6-**23**	2.70	1-6-14	2.80
5	[0.8, 1.0, 0.9]	⑯-㉟-⑩	1-6-**23**	2.70	1-6-17	2.77
6	[1.1, 0.9, 0.8]	㉟-㊳-㉜	8-21-**23**	2.36	7-15-16	2.54
7	[1.0, 0.8, 1.1]	㉟-㊳-㉜	8-21-**23**	2.36	7-15-16	2.54
8	[0.9, 0.8, 1.1]	㉟-㊳-㉜	8-21-**23**	2.36	7-15-16	2.54
9	[1.0, 1.0, 0.8]	⑭-⑰-㉟	12-13-**23**	2.38	7-8-11	2.67
10	[1.0, 1.0, 1.0]	㉟-⑪-⑮	13-14-**23**	2.35	13-14-15	2.43

B. Characteristics of Generator Considered in CFS

Taking the *OC*.10 in Table I as an example, the numbers of lines out of service for CFS1 and CFS2 are the same after the first and second attacks. It is after the third attack that the power system simulated with CFS1 collapse, where attacking transmission line ⑮ resulting in eight more lines out of service

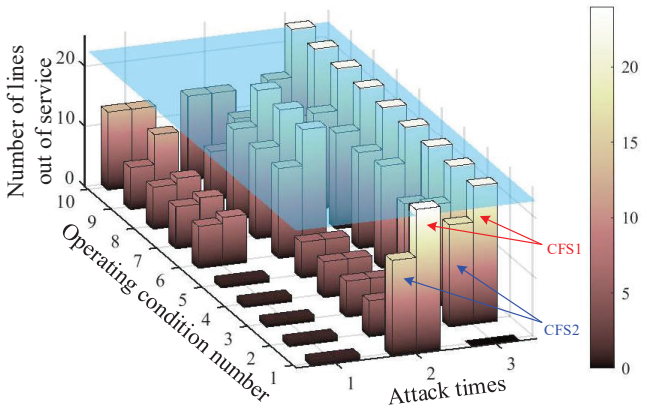


Fig. 4. The number of lines out of service caused by the same attacking sequence with different CFSs under different operating conditions (There are two bars shown at every time attack along the axis of attack. Among them, the left ones represent the number of lines out of service for CFS2, while the right ones represent that for CFS1. The light blue surface parallel to the *xy* plane represents the threshold of the collapse, whose value is 23.)

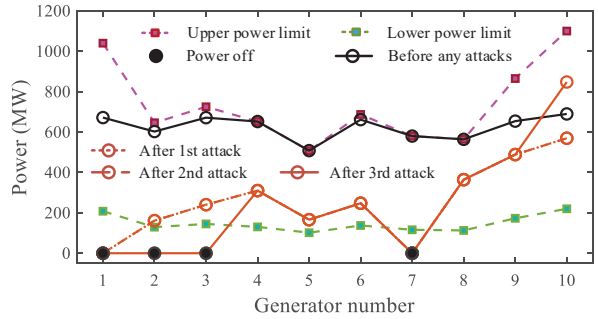
in CFS1 than that in CFS2. Fig. 5 shows the power of each generator under *OC*.10.

The main difference between CFS1 and CFS2 lies in whether the power is less than the lower power limit of the generator. For CFS1 in Fig. 5(a), G_1 and G_7 have been tripped off before the third attack, while G_2 and G_3 also get tripped off after the attack. As for CFS2 in 5(b), G_2, G_3, and G_7 keep connected to the power system but with fairly low power, which could cause the potential damage to the blade of steam turbine of generator and the uneconomical operation. Hence, the resilience assessed with CFS1 is closer to the practical operation of a power system, which considers the governor droop coefficient and lower power limit of generators.

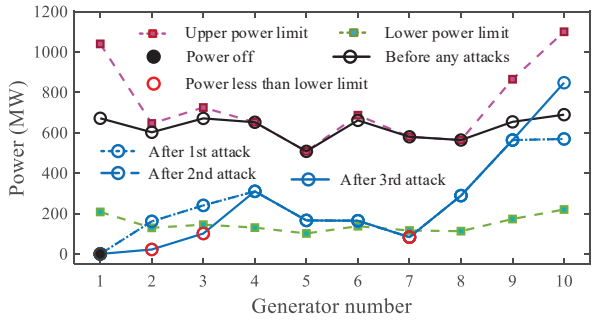
Fig. 6 shows the whole cascading failure process simulated by CFS1 under the sequential attack for *OC*.10. The first attack aims at line ㉟ (connecting bus 21 and bus 22), which is one of the two main power delivery paths of G_2 and G_3. The outage of line ㉟ led to the overload of line ㊳ (connecting bus 23 and 24), which was tripped out later. Then all the power delivery paths of G_6 and G_7 were out of service and the power was far too much for the only load on bus 23 in this sub-grid. As a result, the power system lost about 39.75% power, and G_7 was tripped off. Then the second attack did not result in power loss. Before the third attack, line ⑮ (connecting bus 7 and bus 8) was the only power delivery path for G_2 and G_3. When it was attacked, the left loads in this area were too low to keep anyone of the generators operating economically and safely. It resulted in the tripping off of G_2 and G_3 and about another 1.97% power loss. In the end, over 23 transmission lines were out of service and the power system collapsed. The simulation results indicate that the governor droop coefficient and the lower power limit of generators play an important role in cascading failure simulation.

C. DDQN Agent v.s. Q-Table Agent

In order to verify the capability of the DDQN based agent to assess resilience of power systems under different operating



(a) Simulated by CFS1



(b) Simulated by CFS2

Fig. 5. The power of generators under the sequential attacks for $OC.10$ in 39-bus system

conditions, comparison simulations are conducted with different collapse thresholds for the Q-table based agent and the DDQN based agent, respectively. The collapse thresholds are set to $N_c = 10$, $N_c = 14$, $N_c = 19$ and $N_c = 23$, which are 20%, 30%, 40% and 50% of the total transmission lines in 39-bus power system, respectively. All the simulations are conducted with CFS1, so are the simulations in the next subsection.

The learning parameters of table-based Q-learning agent are set the same as that in [41]. Since the discrete state of the Q-table cannot represent different operating conditions with the same transmission line status, the Q-learning agent is trained and evaluated under each operating condition individually. Then the numbers of attacks required for all operating conditions in evaluation collection are summed to compare with that of DDQN based agent.

As for the DDQN based agent, there are 5 fully connected layers in the deep Q evaluation network. The number of neurons in each layer is set as 28-256-256-256-46. The hyper-parameters of the agent is set as follows: total training steps $N_T = 30000$, memory capacity $M = 10000$, training batch size $B = 256$, learning rate $l_r = 0.001$, model synchronization frequency $f_s = 20$, model evaluation frequency $f_e = 100$, discount factor $\gamma = 0.9$, greedy coefficient $\varepsilon_0 = 1.0$, $\Delta\varepsilon = 0.0001$ and $\varepsilon_f = 0.1$. The DDQN based agent is trained under the training OC collection C_t and evaluated the performance of resilience assessment under evaluation OC collection C_e.

Fig. 7 shows the performance comparison of the Q-table and the DDQN based agent for resilience assessment with different collapse thresholds. It can be seen that the numbers of attacks required by both agents for threshold $N_c = 10$ are almost the same. However, when the threshold increases to $N_c = 23$, the number of attacks required by Q-table agent is almost twice that required by DDQN based agent. That is to say, the number of attacks required to result in collapse under all of C_e increases with the increase of the threshold for power system collapse. What is more, the increase of attacks required by the Q-table based agent is much more obvious than that required by the DDQN based agent.

Meanwhile, it is noticed that the training process of the DDQN agent is relatively more stable than that of the Q-table agent. This is because the increase of threshold leads to the longer MDP of resilience assessment for the power system. As a result, the Q-table needs to explore a larger state-action space for the training process and requires more cache to save and update the Q values corresponding to the explored state-action pairs, which greatly increases the difficulty of the training process convergence and increases the time consumption. In contrast, the DDQN based agent can exploit the transitions more efficiently with the help of experience replay, which is beneficial to the convergence of the training process. Besides, the well trained agent can be used to assess the resilience of power system under different operating conditions.

D. Improved PER v.s. URR

In order to verify the effectiveness of prioritized experience replay for the training of the DDQN agent, comparison simulations are conducted for the DDQN based agent using PER (marked as "PER-DDQN") and DDQN based agent using uniform random replay (URR). The hyper-parameters of PER are set as: the upper limit of priority $\eta = 15$, the priority coefficient $\alpha = 0.6$, and the initial value of IS correction coefficient $\beta_0 = 0.4$, which are recommended in [46]. β increases 2.5×10^{-3} every 100 times of experience replay and the final value is $\beta_0 = 1$. The rest hyper-parameters are the same as that of the DDQN agent. The threshold of power system collapse is $N_c = 23$. Simulations are repeated five times for both PER-DDQN agent and DDQN agent.

Fig. 8 shows the number of attacks required by different resilience assessment methods. The smallest number goes to the traversal method, which is only 29 for ten OCs as summarised in Table I. However, it takes from tens of thousands to millions of trials for the traversal method to find the shortest attack sequence under each OC. It is unacceptable for the resilience assessment relatively larger scale power system with a high collapse threshold. As for the DDQN based agents, the smallest numbers of attacks required by both agents are around 40. The average number of the PER-DDQN agent is about 2 less than that of the DDQN agent at the end of the training process. In addition, the purple dashed line in Fig. 8 represents the smallest number obtained by the DDQN agent, which is obtained by the PER-DDQN agent early at about 12000 training episodes. That means the training of PER-DDQN agent converges faster than DDQN agent. Besides, the performance of the PER-DDQN agent is more stable than that of the DDQN agent at the end of the training process.

Fig. 9 provides the resilience assessment performance comparison of different methods under C_e. As is shown, the

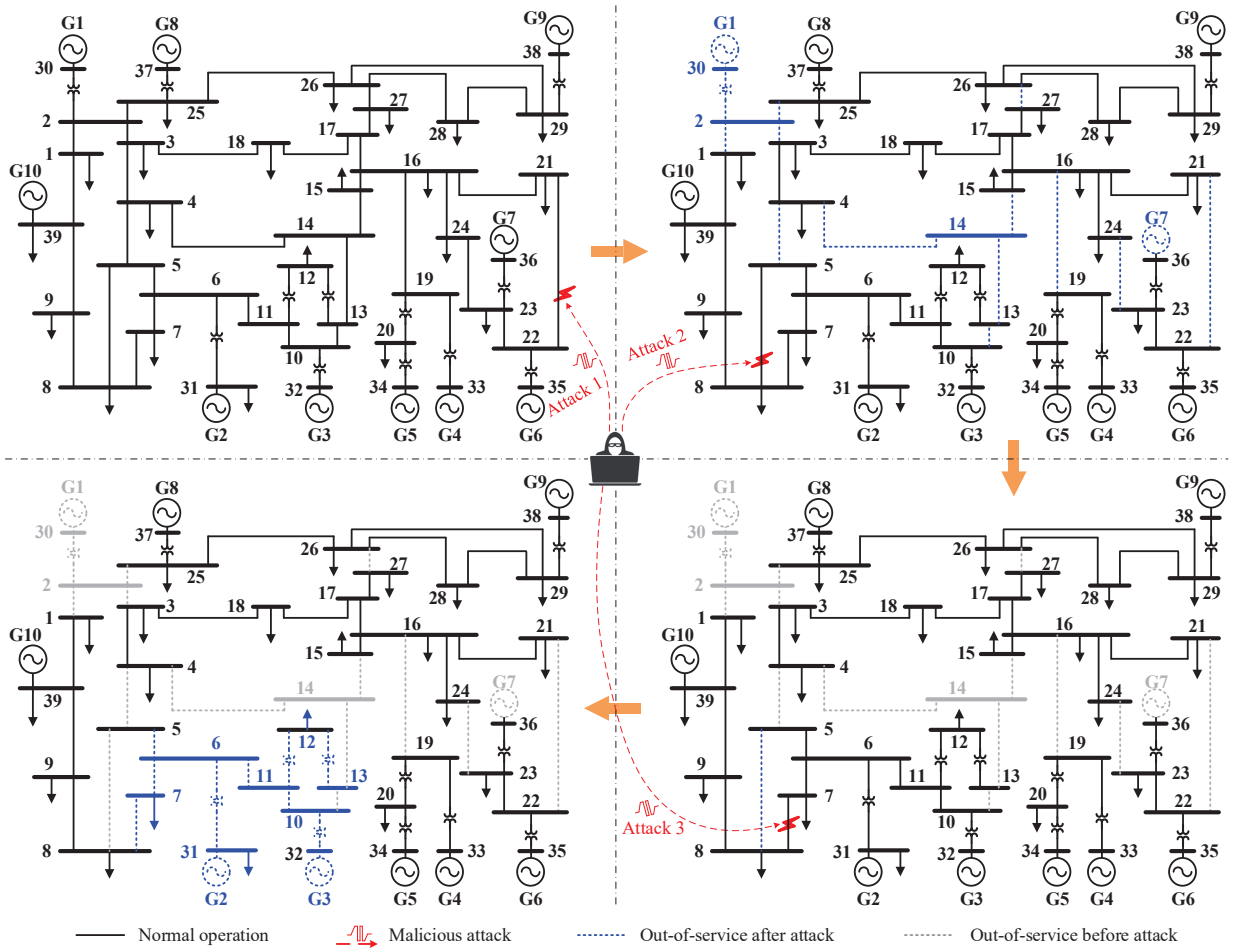


Fig. 6. The collapse process of *OC*.10 under sequential attacks simulated by CFS1

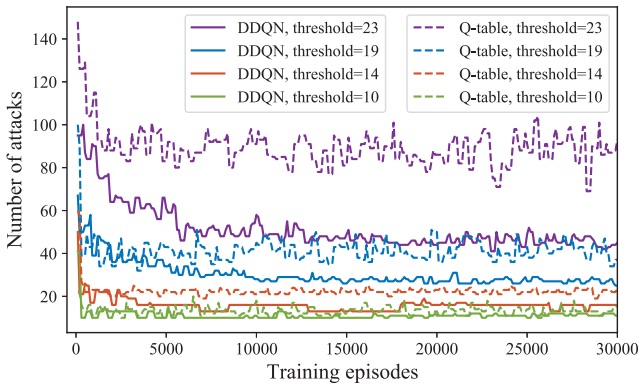


Fig. 7. Performance of DDQN agent compared to Q-table method with different collapse threshold (The numbers of attacks are smoothed by sliding window, which takes the minimum of the three numbers around.)

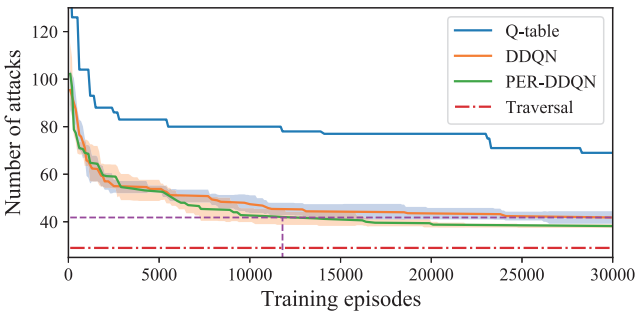


Fig. 8. Performance comparison of different resilience assessment methods with the collapse threshold $N_c = 23$ in 39-bus system

traversal method can make the power system collapse after twice attacks under one of C_e. It can make the power system collapse under all C_e after three times attacks. Among the rest reinforcement learning agents, the resilience assessment performances of DDQN and PER-DDQN are pretty close. PER-DDQN can result in the collapse after 5 times attack under ten OCs, which is two OCs more than that of DDQN. It means that the PER-DDQN agent is slightly better than the DDQN agent. As for the table-based Q-learning agent, there is still only one OC under which the power system would collapse after 6 times attacks.

E. Stochastic Prioritization Replay and the Priority Update

Fig. 10 shows the priority distribution of one batch of transitions replayed by URR and PER. The priorities of the transitions replayed by URR are obviously lower than that of the transitions replayed by PER. In this way, the transitions replayed by PER are relatively more valuable than those replayed by URR for the update of the Q evaluation network.

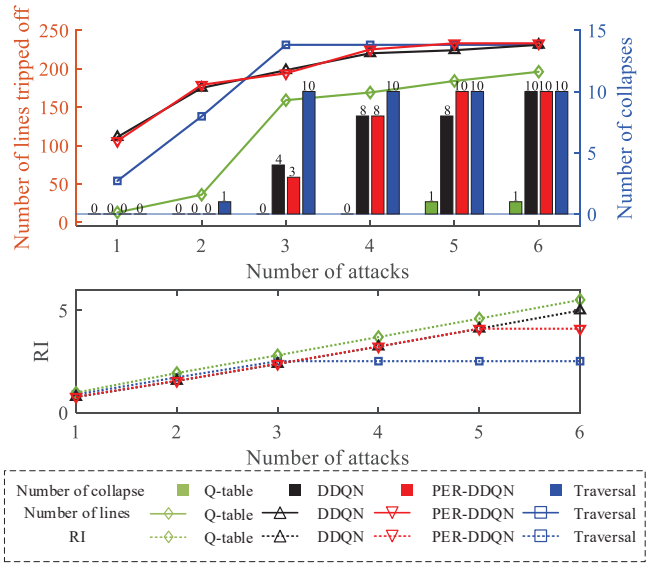


Fig. 9. Evaluation results with different power system resilience assessment methods

The stochastic prioritization replay makes the training process more aggressive, thus making the PER-DDQN converge faster. Also, it can be seen that the priorities of the replayed memories get lower overall after the update of the Q evaluation network. The decline is not too much owing to the average episode reward. In this way, the valuable transitions that come from the episode with a short attacking sequence can keep competitive during the stochastic prioritization replay.

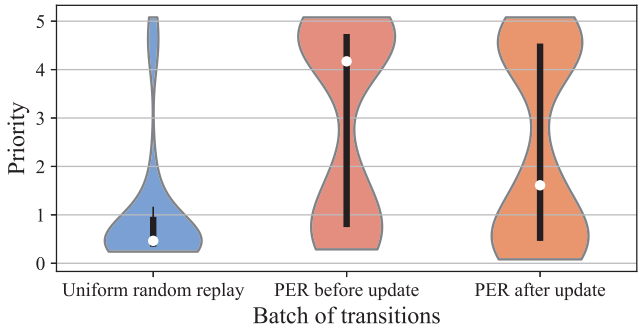


Fig. 10. The priority distribution of the memories sampled by different strategies

F. Importance-Sampling (IS) Weight Correction

Fig. 11 shows the losses calculated from the transitions with and without IS weight correction during the training process. The difference between the two kinds of losses gets larger and larger. There are two main reasons: one is that the principle of transition replacement makes the overall priorities of the transitions in replay buffer higher and higher. It leads to the increase of loss without IS weight correction and is beneficial to the fast convergence. The other goes to that the IS weight correction coefficient β becomes closer to 1 at the last part of the training process, which means that the loss will be fully compensated by IS weight. It shrinks the gradient and makes the learning step smaller. As a result, the resilience assessment performance of PER-DDQN is more stable than that of the DDQN agent at the last part of the training process.

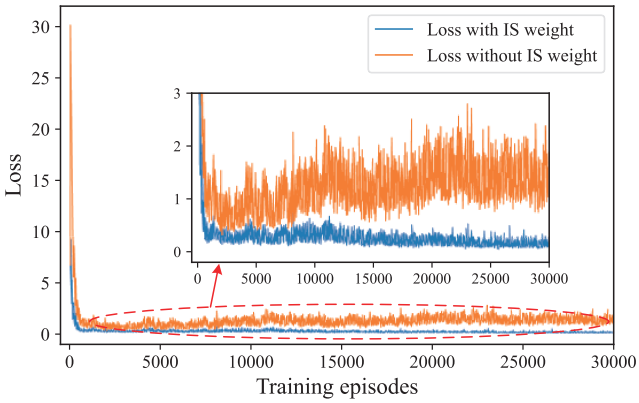


Fig. 11. The loss for the prioritized replayed batch of transitions

G. Transition Replacement in the Improved PER

Fig. 12 shows the priority distribution of the whole transitions in replay buffer during the training process. The priority distribution analyzed at the first time is overall lower than that analyzed at the final time. Specifically, the number of transitions with high priority ($p \geq 3.5$) at the final time is about twice that at the first time. It implies that the proposed transition replacement principle tends to keep the transitions with higher priority in the replay buffer with limited capacity. Together with the prioritization replay, it makes the training process of the PER-DDQN agent more efficient than that of the DDQN agent with URR.

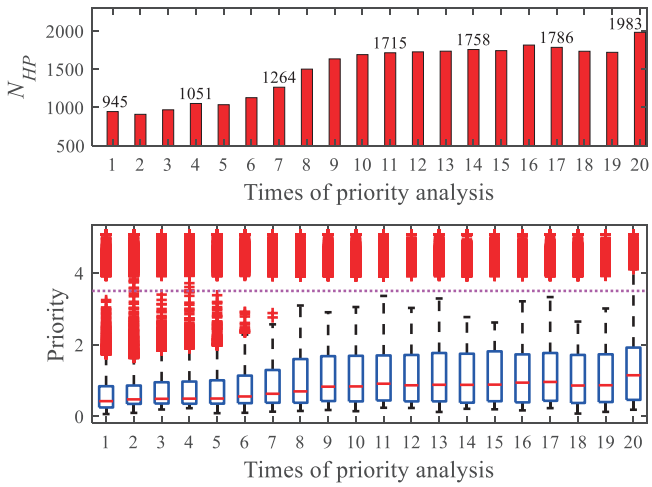


Fig. 12. The priority distribution of the transitions in replay buffer during the training process (N_{HP} represents the number of the transitions with priority higher than 3.5)

H. Effectiveness verification in 118-bus system

The diagram of IEEE 118-bus system is available in [6], which consists of 54 generators, 99 loads, and 186 branches with 11 transformers. Among them, 85 non-transformer branches at voltage level of 138 kV are selected as the critical

lines which are available for the sequential attacks. In order to generate different OCs, the 118-bus systems is divided into three parts. Based on the typical OC, the load and generation level is set from 0.7 to 1.05 with an interval of 0.05, which are 8 choices in total for each part. Among the 512 generated OCs, only 101 OCs meet the $(N-1)$ security constraints for the 85 critical lines. Then, 20 OCs are selected randomly as C_e and the rest OCs are taken as C_t.

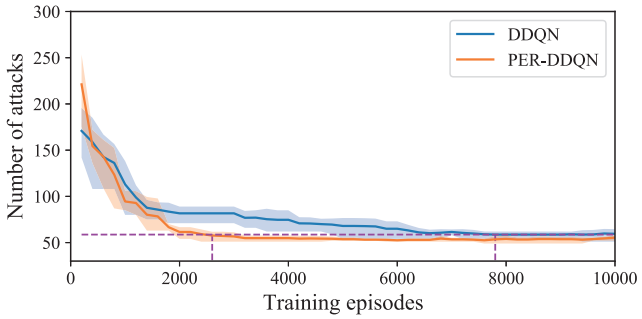


Fig. 13. Performance comparison of different resilience assessment methods with the collapse threshold $N_c = 19$ in 118-bus system (The numbers of attacks required by PER-DDQN and DDQN are smoothed by sliding window, which takes the minimum of the five numbers around.)

The number of neurons in each layer of agent is set as 203-256-256-256-85. The total training step is $N_T = 10000$, and the model evaluation frequency is $f_e = 200$. The rest parameters design of the DDQN based agent in 118-bus system is the same as that in 39-bus system. The threshold of system collapse is $N_c = 19$. Simulations are repeated five times for both PER-DDQN agent and DDQN agent. Fig. 13 shows the comparison results of the performance on C_e. The well-trained PER-DDQN agent needs average 2.65 sequential attacks to make the collapse occur among the 20 evaluation OCs. The results indicate that the training of PER-DDQN agent converges faster than DDQN agent.

I. Effectiveness verification in 300-bus system

The IEEE 300-bus system consists of 69 generators, 199 loads, and 411 branches with 107 transformers. Among them, 60 non-transformer branches at relatively lower voltage levels of 115 *kV* and 230 *kV* are selected as the critical lines. Similarly, the 300-bus systems is divided into three parts. The load and generation level is set from 0.75 to 1.1 with an interval of 0.05, which are 8 choices in total for each part. Among the 512 generated OCs, only 309 OCs meet the $(N-1)$ security constraints for the 60 critical lines. Then, 80 OCs are selected randomly as C_e and the rest OCs are taken as C_t.

The number of neurons in each layer of agent is set as 258-256-256-256-60. The parameters design of the DDQN based agent in 300-bus system is the same as that in 118-bus system. The threshold of system collapse is $N_c = 21$. Simulations are repeated five times for both PER-DDQN agent and DDQN agent. Fig. 14 shows the comparison results of the performance on C_e. The well-trained PER-DDQN agent needs average 3.32 sequential attacks to make the collapse occur among the 80 evaluation OCs. The results also indicate that the faster convergence process of the training for PER-DDQN agent. Besides, it reflects that the performance of PER-DDQN agent is more stable than that of the DDQN agent.

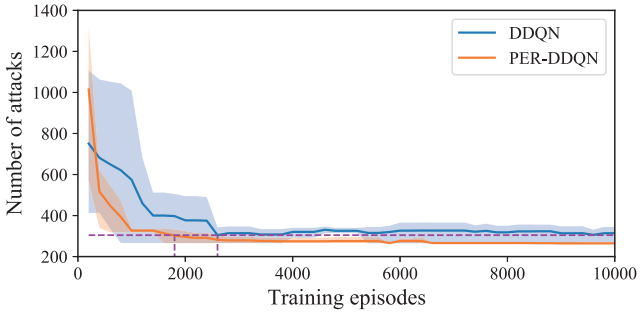


Fig. 14. Performance comparison of different resilience assessment methods with the collapse threshold $N_c = 21$ in 300-bus system (The numbers of attacks required by PER-DDQN and DDQN are smoothed by sliding window, which takes the minimum of the five numbers around.)

V. Conclusion

This paper proposes a DDQN based resilience assessment method for power systems under sequential attacks. A DC power flow based cascading failure simulator is proposed to simulate the topology and power flow changes, which can take the governor droop coefficient and lower bound generation of generators into consideration. Besides, an index based on the number of attacks is proposed for the resilience assessment, which reflects the capability of the power system to tolerate attacks. Then a DDQN based agent is proposed to conduct the resilience assessment under different operating conditions. In addition, the improved PER technique is developed for the training of the DDQN based agent, which scores the priority of transition by the combination of TD-error and average episode reward.

Simulation results in the 39-bus, 118-bus and 300-bus power systems verify the superiority of the proposed resilience assessment method. The resilience assessed with the proposed CFS is more prudential than that without the sufficient consideration of generator characteristics. The proposed DDQN based agent can assess the resilience of power systems under different operating conditions. Compared to uniform random replay technique, the improved PER technique brings faster convergence and more stable resilience assessment performance in the last part of the training process.

Future work will focus on the development of the deep reinforcement learning based defender to protect the power network from collapse under malicious attacks.

References

[1] S. Paul, F. Ding, K. Utkarsh, W. Liu, M. J. O'Malley, and J. Barnett, "On vulnerability and resilience of cyber-physical power systems: A review," *IEEE Syst. J.*, doi: 10.1109/JSYST.2021.3123904.

[2] A. Clark, and S. Zonouz, "Cyber-physical resilience: Definition and assessment metric," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1671-1684, Mar. 2019.

[3] E. Bompard, C. Gao, R. Napoli, A. Russo, M. Masera, and A. Stefanini, "Risk assessment of malicious attacks against power systems," *IEEE Trans. Syst., Man, A, Syst. Humans*, vol. 39, no. 5, pp. 1074-1085, Sep. 2009.

[4] E. I. Bilis, W. Kröger and C. Nan, "Performance of electric power systems under physical malicious attacks," *IEEE Syst. J.*, vol. 7, no. 4, pp. 854-865, Dec. 2013.

[5] K. Lai, M. Illindala, and K. Subramaniam, "A tri-level optimization model to mitigate coordinated attacks on electric power systems in a cyber-physical environment," *Appl. Energy*, vol. 235, no. 2019, pp. 204-218, Feb. 2019.

[6] P. Cuffe, "A comparison of malicious interdiction strategies against electrical networks," *IEEE Trans. Emerg. Sel. Topics Circuits, Syst.*, vol. 7, no. 2, pp. 205-217, Jun. 2017.

[7] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317-3318, Jul. 2017.

[8] P. Gautam, P. Piya, and R. Karki, "Resilience assessment of distribution systems integrated with distributed energy resources," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 338-348, Jan. 2021.

[9] L. Xu, Q. Guo, Y. Sheng, S.M. Muyeen, and H. Sun, "On the resilience of modern power systems: A comprehensive review from the cyber-physical perspective," *Renew. Sust. Energ. Rev.*, vol. 152, no. 2021, pp. 111642, Dec. 2021.

[10] A. Beiranvand and P. Cuffe, "Negative results on deploying distributed series reactance devices to improve power system robustness against cascading failures," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5210-5221, Mar. 2021.

[11] F. H. Jufri, V. Widiputra, and J. Jung, "State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies," *Appl. Energy*, vol. 239, no. 2019, pp. 1049-1065, Apr. 2019.

[12] R. Arghandeh, A. V. Meier, L. Mehrmanesh, and L. Mili, "On the definition of cyber-physical resilience in power systems," *Renew. Sust. Energ. Rev.*, vol. 58, no. 2016, pp. 1060-1069, May 2016.

[13] M. Bao, Y. Ding, M. Sang, D. Li, C. Shao, and J. Yan, "Modeling and evaluating nodal resilience of multi-energy systems under windstorms," *Appl. Energy*, vol. 270, no. 2020, pp. 115136, Jul. 2020.

[14] S. Espinoza, M. Panteli, P. Mancarella, and H. Rudnick, "Multi-phase assessment and adaptation of power systems resilience to natural hazards," *Electr. Power Syst. Res.*, vol. 136, no. 2016, pp. 352-361, Jul. 2016.

[15] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1294-1305, Jul. 2013.

[16] X. Liu and Z. Li, "Local topology attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2617-2626, Nov. 2017.

[17] S. Lakshminarayana, J. S. Karachiwala, T. Z. Teng, R. Tan, and D. K. Y. Yau, "Performance and resilience of cyber-Physical control systems with reactive attack mitigation," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6640 - 6654, Nov. 2019.

[18] L. Che, X. Liu and Z. Li, "Mitigating false data attacks induced overloads using a corrective dispatch scheme," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3081-3091, May 2019.

[19] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vo. 27, no. 8, pp. 1773-1786, Aug. 2016.

[20] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "A framework for cyber-topology attacks: Line-switching and new attack scenarios," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1704-1712, Mar. 2019.

[21] A. Street, F. Oliveira, and J. M. Arroyo, "Contingency-constrained unit commitment with n-k security criterion: A robust optimization approach," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1581-1590, Aug. 2011.

[22] J. Yan, Y. Zhu, H. He, and Y. Sun, "Multi-contingency cascading analysis of smart grid based on self-organizing map," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, pp. 646-656, Apr. 2013.

[23] X. Gao, C. Pu, X. Chen, and L. Li, "Vulnerability assessment of power grids against cost-constrained hybrid attacks," *IEEE Trans. Circuits Syst. II-Express Briefs*, vol. 68, no. 4, pp. 1477-1481, Apr. 2021.

[24] Y. Zhu, J. Yan, Y. Tang, Y. Sun, and H. He, "Resilience analysis of power grids under the sequential attack," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2340-2354, Dec. 2014.

[25] L. Che, X. Liu, Z. Li, and Y. Wen "False data injection attacks induced sequential outages in power systems," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1513-1523, Mar. 2019.

[26] T. Ding, M. Qu, X. Wu, B. Qin, Y. Yang and F. Blaabjerg, "Defense strategy for resilient shipboard power systems considering sequential attacks," *IEEE Trans. Inf. Forensics Security*, vol. 15, no. 12, pp. 3443-3453, Dec. 2019.

[27] Y. Huang, J. Wu, C. K. Tse, and Z. Zheng, "Sequential attacker-defender game on complex networks considering the cascading failure process," *IEEE Trans. Comput. Social Syst.*, early access, Aug. 10, 2021, doi: 10.1109/TCSS.2021.3099718.

[28] M. H. Athari, and Z. Wang, "Impacts of wind power uncertainty on grid vulnerability to cascading overload failures," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 128-137, Jan. 2018.

[29] J. Yan, Y. Tang, H. He, and Y. Sun, "Cascading failure analysis with DC power flow model and transient stability analysis," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 285-297, Jan. 2015.

[30] N. Alguacil, A. Delgadillo, and J. M. Arroyo, "A trilevel programming approach for electric grid defense planning," *Comput. Oper. Res.*, vol. 41, no. 2014, pp. 282-290, Jan. 2014.

[31] P. Rezaei and P. D. Hines, "Changes in cascading failure risk with generator dispatch method and system load level," *in Proc. IEEE PES TD Conf. Expo.*, Apr. 2014, pp. 1-5.

[32] M. J. Eppstein, P. D. H. Hines, "A 'Random Chemistry' algorithm for identifying collections of multiple contingencies that initiate cascading failure," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1698-1705, Aug. 2012.

[33] M. Noebels, R. Preece, and M. Panteli, "AC cascading failure model for resilience analysis in power networks," *IEEE Syst. J.*, early access, pp. 1-12, Dec. 2020; doi: 10.1109/ JSYST.2020.3037400.

[34] G. Wu, M. Li, and Z. S. Li, "A Stochastic modeling approach for cascading failures in cyberphysical power systems," *IEEE Syst. J.*, early access, pp. 1-12, Apr. 2021; doi: 10.1109/JSYST.2021.3070503.

[35] X. Gao, M. Peng, C. K. Tse, and H. Zhang, "A stochastic model of cascading failure dynamics in cyber-physical power systems," *IEEE Syst. J.*, vol. 14, no. 3, pp. 4626-4637, Sep. 2020.

[36] E. Bompard, A. Estebsari, T. Huang, and G. Fulli, "A framework for analyzing cascading failure in large interconnected power systems: A post-contingency evolution simulator," *Int. J. Electr. Power Energy Syst.*, vol. 81, no. 2016, pp. 12-21, Oct. 2016.

[37] R. Rocchetta, E. Zio, and E. Patelli, "A power-flow emulator approach for resilience assessment of repairable power grids subject to weather-induced failures and data deficiency," *Appl. Energy*, vol. 210, no. 2018, pp. 339-350, Jan. 2018.

[38] M. Panteli, P. Mancarella, "Modeling and Evaluating the Resilience of Critical Electrical Power Infrastructure to Extreme Weather Events," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1733-1742, Sep. 2017.

[39] V. Fathi, J. Arabneydi, and A. G. Aghdam, "Reinforcement learning in linear quadratic deep structured teams: Global convergence of policy gradient methods," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2020, pp. 4927-4932.

[40] D. M. Casas-Velasco, O. M. C. Rendon, and N. L. S. da Fonseca, "DRSIR: A deep reinforcement learning approach for routing in software-defined networking," *IEEE Trans. Netw. Serv. Manag.*, early access, Dec. 2021; doi: 10.1109/TNSM.2021.3132491.

[41] J. Yan, H. He, X. Zhong, and Y. Tang, "Q-Learning-based vulnerability analysis of smart grid against sequential topology attacks," *IEEE Trans. Inf. Forensic. Secur.*, vol. 12, no. 1, pp. 200-210, Jan. 2017.

[42] H. V. Hasselt, A. Guez, D. Silver, "Deep reinforcement learning with double q-learning," *AAAI 2016*, vol. 30, no. 1, pp. 2094-2100, Feb. 2016.

[43] Z. Wang, H. He, Z. Wan and Y. Sun, "Coordinated topology attacks in smart grid using deep reinforcement learning," *IEEE Trans. Ind. Inform.*, vol. 17, no. 2, pp. 1407-1415, Feb. 2021.

[44] Y. Koc, T. Verma, N. A. M. Araujo, and M. Warnier, "MATCASC: A tool to analyse cascading line outages in power grids," *in Proc. 2013 IEEE Int. Workshop Inteligent Energy Syst.*, pp. 143-148, Nov. 2013.

[45] M. Dorokhova, Y. Martinson, C. Ballif, and N. Wyrsch, "Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation," *Appl. Energy*, vol. 301, no. 2021, pp. 117504, Nov. 2021.

[46] T. Schaul, J. Quan, I. Antonoglou and D. Silver, "Prioritized experience replay," *ICLR 2015*, arXiv preprint arXiv:1511.05952.

[47] J. Li, T. Yu, X. Zhang, F. Li, D. Lin, and H. Zhu, "Efficient experience replay based deep deterministic policy gradient for AGC dispatch in integrated energy system," *Appl. Energy*, vol. 285, no. 2021, pp. 116386, Mar. 2021.

[48] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12-19, Feb. 2011.

Lingkang Zeng received the B.S. degree in 2016 in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, where he is currently working toward the Ph.D. degree in electrical engineering.

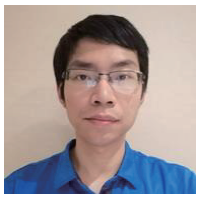
He was also a Visiting Student Researcher with the University of Rhode Island, Kingston, RI, USA, from 2019 to 2020. His current research interests include deep reinforcement learning, power system security assessment and stability control.

Wei Yao (M'13-SM'17) received the B.S. and Ph.D. degrees in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2004 and 2010, respectively.

He was a Post-Doctoral Researcher with the Department of Power Engineering, HUST, from 2010 to 2012 and a Postdoctoral Research Associate with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K., from 2012 to 2014. Currently, he has been a Professor with the School of Electrical and Electronics Engineering, HUST, Wuhan, China. His current research interests include power system stability analysis and control, renewable energy, HVDC and DC Grid, and application of artificial intelligence in Smart Grid.

Hang Shuai (S'17-M'19) received the B.Eng. degree from Wuhan Institute of Technology (WIT), Wuhan, China, in 2013, and the Ph.D. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019.

He was also a visiting student researcher with the University of Rhode Island (URI), Kingston, RI, USA, from 2018 to 2019. He was a Postdoctoral Researcher with the University of Rhode Island from 2019 to 2020. Currently, he is a Research Scientist with the University of Tennessee, Knoxville, TN, USA. His research interests include deep reinforcement learning, microgrid optimization, bulk power system resilience.



Yue Zhou (M'13) received his BSc, MSc and Ph.D. degrees in Electrical Engineering from Tianjin University, China, in 2011, 2016 and 2016, respectively.

He is the Lecturer in Cyber Physical Systems at School of Engineering of Cardiff University, Wales, UK. His research interests include demand response, peer-to-peer energy trading and cyber physical systems. Dr Zhou is a Managing Editor of Applied Energy, and an Associate Editor of IET Energy Systems Integration, IET Renewable Power Generation and Frontiers in Energy Research. He is the Chair of CIGRE UK Next Generation Network (NGN) Committee. He is also a Committee Member of IEEE PES UK & Ireland Chapter.

Xiaomeng Ai (S'11-M'17) received the B.S. degree in mathematics and applied mathematics and the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2008 and 2014, respectively.

He is currently an Associate Professor with HUST. His research interests include robust optimization theory in power system, renewable energy integration, and integrated energy market.

Jinyu Wen (M'10) received the B.S. and Ph.D. degrees in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1992 and 1998, respectively.

He was a visiting student from 1996 to 1997 and a Research Fellow from 2002 to 2003 with the University of Liverpool, Liverpool, U.K., and a Senior Visiting Researcher with the University of Texas at Arlington, Arlington, USA, in 2010. From 1998 to 2002, he was the Director Engineer with XJ Electric Company Ltd., China. In 2003, he joined HUST, where he is currently a Professor with the School of Electrical and Electronics Engineering. His current research interests include renewable energy integration, energy storage, multi-terminal HVDC, and power system operation and control.