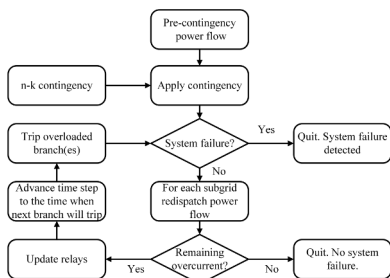


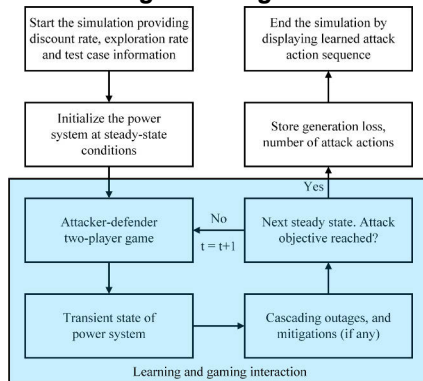
Abstract

Because of the increasing number of heterogeneous devices connected to electric power grid, the attack surface increases the threat actors. Game theory and machine learning are being used to study power system failures caused by external manipulation. Most existing works in the literature focus on the one-shot process of attacks and fail to show the dynamic evolution of the defense strategy. In this paper, we focus on an adversarial, multistage, sequential game between the adversaries of the smart electric power transmission and distribution system. We study the impact of the exploration rate and the convergence of the attack strategies (sequences of action that create large-scale blackouts based on the system capacity) based on the reinforcement learning approach. We also illustrate how the learned attack actions disrupt the normal operation of the grid by creating transmission line outages, bus voltage violations, and generation loss. This simulation studies are conducted on IEEE 9- and 39-bus systems. The results show the improvement of the defense strategy through the learning process. The results also prove the feasibility of the learned attack actions by replicating the disturbances created in simulated power systems.

Threat Model



Multistage Gaming Framework



Reward Design for Q-Learning

The reward for the Q-learning agent is assigned as follows:

$$R^A(s, a, d) = \begin{cases} +1, & \text{if } N_0 \geq N_\theta \text{ and } k < N_\theta \\ 0, & \text{otherwise} \end{cases}$$

$$R^D(s, a, d) = \begin{cases} -1, & \text{if } N_0 \geq N_\theta \text{ and } k \geq N_\theta \\ 0, & \text{otherwise} \end{cases}$$

N_θ : attack objective

N_0 : number of transmission line failures

k : number of actions taken by the attacker.

Settings

Table I: Parameter information for IEEE 39 bus system.

Parameter	Value
Total number of branches, N	46
Total generation	6150 MW
Attack objective, N_θ	30% of line outage (minimum 14)
Defender's action set	(1, 2, 3)
Discount factor, γ	0.9
Maximum iteration per episode	100
Total number of episodes	8000
Total number of run	100
Exploration probability, ϵ	0.8
Final epsilon, ϵ_f	0.0004
Epsilon divided	Every 100 episode
Epsilon divided by	1.1

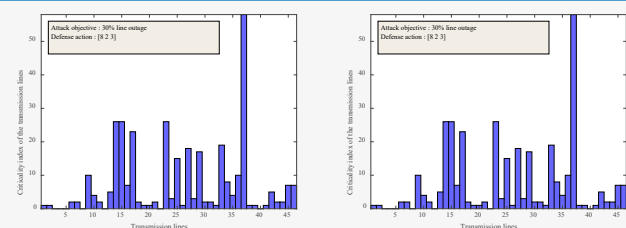
Table III: Number of episodes to be explored with different exploration rates for IEEE 39 bus system

Exploration rate	Numbers of episodes to be explored	% of convergence to optimality
20%	1,600 (out of 8,000)	5%
40%	3,200 (out of 8,000)	75%
80%	6,400 (out of 8,000)	95%

Case Study 1

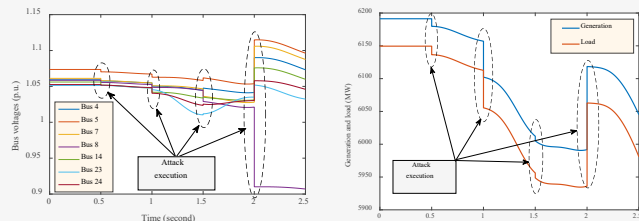
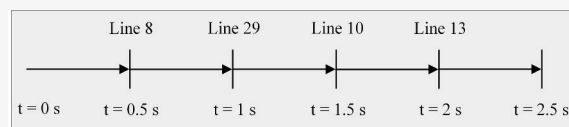
Table II: Number of episodes to be explored with different exploration rates for IEEE 9 bus system

Exploration rate	Numbers of episodes to be explored	% of convergence to optimality
20%	1,000 (out of 5,000)	15%
40%	2,000 (out of 5,000)	30%
80%	4,000 (out of 5,000)	100%



Criticality index of the transmission lines of IEEE 39-bus system to be selected as the attack actions by the attacker when the defender defends transmission line (a) {8,2,3}. The number of unique optimal policies here is 82, and (b) {8, 27, 37}. The number of unique optimal policies here is 19.

Case Study 2



(a) Voltages (p.u.) at the buses connecting the target transmission lines and (b) change in generation and load (in MW) because of the attacks in the IEEE 39-bus system