A Study of Linear Programming and Reinforcement Learning for One-Shot Game in Smart Grid Security

Shuva Paul and Zhen Ni
Department of Electrical Engineering and Computer Science
South Dakota State University, Brookings, SD 57006
United States
Email: {shuva.paul, zhen.ni}@sdstate.edu

***Abstract*—Smart grid attacks can be applied on a single component or multiple components. The corresponding defense strategies are totally different. In this paper, we investigate the solutions (e.g., linear programming and reinforcement learning) for one-shot game between the attacker and defender in smart power systems. We designed one-shot game with multi-line-switching attack and solved it using linear programming. We also designed the game with single-line-switching attack and solved it using reinforcement learning. The pay-off and utility/reward of the game is calculated based on the generation loss due to initiated attack by the attacker. Defender's defense action is considered while evaluating the pay-off from attacker's and defender's action. The linear programming based solution gives the probability of choosing best attack actions against different defense actions. The reinforcement learning based solution gives the optimal action to take under selected defense action. The proposed game is demonstrated on 6 bus system and IEEE 30 bus system and optimal solutions are analyzed.**

***Index Terms*—Game theory, smart grid security, reinforcement learning, cascading failures, linear programming.**

I. Introduction

Smart electric power grid is inter-connected within various nations and regions. Large scale generation units, various residential and industrial loads, transmission and distribution systems, distributed energy sources etc. are making the power system meeting the demand of electrical power in this age. Modernization of the power grid (cyber physical power system) has brought flexible and efficient generation, transmission and distribution of electrical power along with high exposure to severe vulnerabilities [1]–[4]. Events in the power system like transmission line outages, generator outages, load variation and many more are making the future power system challenging to keep safe, controlled and monitored. Additionally, combining the traditional power grid with the cyber operation, control and monitoring is making the power system vulnerable to several threats including cyber attack in the operating stations, transmission line outages, false data injections, malware injections etc. Cascading failures, cyber-physical system security, hardware/human-in-the-loop etc. are significant areas in the smart grid security research for modern smart power system [5].

Game theory is an analytical tool to analyze the complex interactions between dependent/independent rational players with a set of mathematical rules and framework. System operators in the power system usually monitors the system's health and take actions accordingly. In case of contingencies, operators take corrective actions (i.e. remedial actions) to restore the system back to normal operating condition [6]. Attacker attacks the power system (physically or/and virtually) with the intention of harming/sabotizing the power system and create electrical, financial and social hazards. Game theory is used to explain complex interactions between the attacker and the defender (power system operator) in the smart power system successfully from various points of view [7]. Use of game theory is also discovering many emergent areas of vulnerabilities in smart grid security which needs to be explored more. In case of defending, the power system operators and the power system protection schemes monitor the system's health by observing the key parameters (e.g. voltage, power (real and reactive), frequency etc.). After observing the key parameters, the defender takes corrective actions in case of any emergency situations like generator outages, transmission line failures, overloading of the transmission lines, relay malfunctioning etc. Incidents like northeast power grid blackout [8], cyberattack on ukraine power grid [9], Stuxnet's attack in iran's nuclear program[10] etc. proves the necessity of improving the security of smart power system. These improvements can be done by understanding the strategies of cyber attackers by finding the vulnerabilities of power systems.

Existing research in power system security using game theory gives the benefit of understanding the attacker-defender interactions in the power system. This improves the power system protection schemes by allowing the authorities (operators, utility companies and governments) to find the breaches/ weak areas in the power system and protect them. Static and dynamic games are two branches in game theory. Static games are formulated to observe the interactions between the players for one action. In dynamic game, interactions are observed between the players for multiple interdependent actions [11]. In [12], game theoretic environment is designed and staged to analyze the cyber switching attack by observing voltage angles and power flows for all the generators. A two-player zero-sum game is introduced between attacker and defender to evaluate the game equilibrium of defense mechanisms under network configurations in [13]. In [14] and [15] static game theory is used to identify the vulnerable and critical components of a smart electric power system considering attacker can conduct

only one action. In [16], power system measurements are attacked in a static game to investigate the interactions between attacker and defender.

Inspired by the gaming in smart grid security researches, we are proposing solutions for the one-shot game in smart electric power grid. Both single line outage and multiple transmission line outages are considered for attacker's action set. The linear programming based proposed approach for the one-shot game will give multiple solutions for different attack-defense action sets with probabilities to be executed using multi-line-switching attack. A pay-off matrix is formulated to solve the game. Generation loss due to the line switching is considered as the pay-offs of the game matrix. The solution using reinforcement learning will also give the optimal attack action from attacker's perspective by using single line-switching attack. From reinforcement learning based solution we can see that, defender's action can be changed from attacker's previous action history. This change will strengthen the security of power system by identifying and protecting critical elements, reducing generation and financial loss.

The rest of the papers are organized as follows: section II describes the formulation of the game problem, calculation of generation loss, selection of targets and attack matrices etc. Gaming solution between attacker and defender using linear programming and reinforcement learning is explained in section III. Simulation results are discussed in section IV. Finally section V gives the summary of this paper's conclusion.

II. Problem formulation and implementation

In this section, the game between attacker and defender is formulated and solved following two different approaches. Linear programming is used to solve the game for multi-line-switching attack. Reinforcement learning is used to solve the game for single-line-switching attack. Problem formulation for two-player zero-sum game between the attacker and the defender is shown using game matrix for linear programming based solution. To formulate the game with game matrix, in this section, calculation of the generation loss is explained briefly. Load ranking of the buses is used here to identify the most critical buses and the transmission lines connected to them for any typical power system. For reinforcement learning based solution, value iteration method is used. After value iteration, optimal attack target is achieved by fighting against a static defender. To rank the buses according to their loads, transmission lines connected to individual buses are identified. Then, transmission lines connected to the individual buses are used as the targets to trigger line-switching attacks. For individual buses, generation losses are calculated using a modified dc cascaded failure simulator [17] [18]. Generation loss is also used as the reward for reinforcement learning based problem formulation and solution.

A. Benchmark model

The solution of the game is proposed using two different approaches. To demonstrate the proposed solution of the two-player, zero-sum attacker-defender game using linear programming, IEEE 30 bus system is used. 6 bus system is used to demonstrate the proposed solution of the game using reinforcement learning.

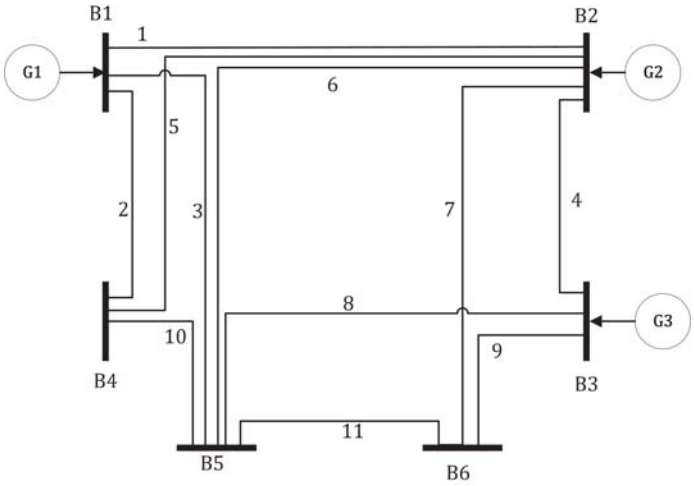

Figure 1: One-line diagram of 6 bus system. Topological information of this test system is used to create attack and defense vectors and evaluate the attack-defense strategy.

Figure 1 shows the one-line diagram of 6 bus system. The topological information (such as the connection between the buses, branch indices, maximum power flow between the buses etc.) is used in the simulation to calculate the generation loss. Transmission line indices are also used to represent the attacks in the power system test cases.

B. Calculation of generation loss

To solve the game between the attacker and the defender in smart grid security, calculation of generation loss is required both in linear programming and reinforcement learning based solution. To calculate the generation loss, a modified DC cascaded failure simulator named DCSIMSEP is adopted [17] [18] [19]. In this paper, modified DC cascaded failure simulator is capable of taking a vector of transmission lines as input and gives the generation losses due to the attacks in these transmission lines. Simultaneous and sequential attacks are types of attack strategies in the smart grid security. Simultaneous attack represents attacking multiple targets at the same time while sequential attack represents attacking multiple targets in a time sequence (one after another) [20][21]. Simultaneous attack strategy is adopted to calculate generation loss in this paper for solving the game using linear programming. To solve one-shot game using reinforcement learning, single-line-switching attack at a time is considered.

Algorithm 1 shows the process of calculation of generation loss using the simulator. In order to calculate the generation loss, indices of the target branches are given to the system. Then, the simulator initializes the power flow. It starts the simulation with measuring the pre-contingency power flow.

Algorithm 1: Calculation of generation loss matrix

Input : Case name, attack matrix containing transmission line indices
Output: Cascaded outages, total generation losses
Result: Rebalanced generation loss matrix

```
Initialization;
for Given test case do
    Load the test case;
    Extract the attack matrix;
    for Given attack matrix do
        Run power flow;
        Switch the branches from attack matrix;
        Divide into sub-grids according to the overloads;
        Redispatching the power flow;
        Update the relay settings;
        if There are overloads then
            Trip the branches according to updated settings;
            Check for the overloads again;
        else
            Calculate total generation loss;
        end
    end
    Display the generation loss matrix;
end
```

The power flow is used to calculate the system status and measure the overloads. These overloads result in tripping some branches as a consequence of cascading failure due to overcurrent. Then redispatching and recalculating of power flow is done in the system. It gives the separation of the grid into subgrids and redispatching the power flow in the subgrids. The generators in the system are then ramped up or down to rebalance the power flow within the range of $P_{\max}$ and $P_{\min}$. After redispatching if there is any surplus of generation in the subgrids, the system trips the generators in the subgrids one by one. After redistributing the power flow or redispatching the generators, if there is a surplus of generation, the generators are tripped down. Usually, the generators are tripped sequentially according to the generation capacity, from small to large.

$$R = (\sum_{g \in G} P_g - \sum_{d \in D} P_d > 0) \tag{1}$$

where P_g is the generated power, P_d is the load demand, G is the set of generator buses and D is the set of load buses. The tripping of the generators continues till $R \leq 0$.

Usually in the surplus cases, instead of tripping the generators, ramping the generation can be proved to be a slow process. If automatic generation control (AGC) fails to fix the frequency error, a few generators will trip because of the overspeed relays. Even if after this, $R < 0$ load shedding occurs by multiplication with a scalar factor λ. λ is defined as

$$\lambda = \frac{\sum_{g \in G} P_g}{\sum_{d \in D} P_d}. \tag{2}$$

After that, DC power flow is simulated in the simulator. Then, the relay settings are updated. Usually, time delayed overcurrent relays are used in the simulator to identify the branches to be tripped due to overcurrent. There is an overcurrent threshold which is fixed by the system operator termed as $\bar{o_j}$. For branch j, for the power flow of f_j and flow limit of $\bar{f_j}$, the outage happens when concurrent overload o_j crosses the limit $\bar{o_j}$. This concurrent overload is calculated from:

$$\Delta o_j(t, \Delta t) = \begin{cases} \int_t^{t+\Delta t} (f_j(t) - \bar{f_j}) dt & \text{if} f_j(t) > \bar{f_j} \\ 0 & otherwise \end{cases} \tag{3}$$

The simulator finds the minimum time for failing the next branch. This time is denoted as ΔT. Then, the time is advanced or updated with the addition of ΔT. Then if the relay trips due to overcurrent, it will switch the online branches to offline. Thus, the generation loss is calculated. These generation losses are used both in the solution of the game using linear programming and reinforcement learning.

C. Attack matrix, selection of targets, generation losses and game matrix

Attack matrix is required for the game. This matrix contains the branches associated with the buses of a typical power system. Individual rows represent the buses and columns represent the branches associated with the buses. Table I shows the generation loss for attacks in the transmission lines connected to individual buses. After calculation of generation loss, the target buses are selected based on the criticality. The criticality is determined based on the generation loss due to switching the lines connected to the buses. The more generation loss occurs due to attacking one bus, the more it is critical than others. For example, from Table I, bus 6 connects transmission lines $6, 7, 9, 10, 11, 12$ and 41. Triggering line-switching attack in these transmission lines will cause generation loss of 70.49 MW. Similarly, bus 9 connects transmission lines $11, 13$ and 14. Triggering line-switching attack in these transmission lines will cause no generation loss. Because these transmission lines are not connected to the generator buses. It means, branches associated with bus 6 is more critical than branches associated with bus 9 in selecting the target to attack.

Table I: Generation losses for the attack matrix of IEEE 30 bus system. These generation losses are the pre-calculation to identify the branches connected with individual buses with the highest effect on the system.

Bus	Branches	Loss	Bus	Branches	Loss
1	[1 2]	10.79	16	[19 21]	3.5
2	[1 3 5 6]	48.22	17	[21 26]	9
3	[2 4]	2.4	18	[22 23]	3.2
4	[3 4 7 15]	7.6	19	[23 24]	9.5
5	[5 8]	0	20	[24 25]	2.2
6	[6 7 9 10 11 12 41]	70.49	21	[27 29]	17.5
7	[8 9]	22.8	22	[28 29 31]	7.39
8	[10 40]	30	23	[30 32]	3.95
9	[11 13 14]	0	24	[31 32 33]	8.7
10	[12 14 25 26 27 28]	5.8	25	[33 34 35]	3.5
11	[13]	0	26	[34]	3.5
12	[15 16 17 18 19]	29.70	27	[35 36 37 38]	13
13	[16]	22.25	28	[34 40 41]	0
14	[17 20]	6.2	29	[37 39]	2.4
15	[18 20 22 30]	8.2	30	[38 39]	10.6

From the generation losses showed in Table I, we select buses $8, 7, 2, 12, 6$ and 13 as targets for attacking and defending. We name them as $z1, z6, z2, z3, z4$ and $z5$ respectively. Then we make the combination matrix taking two targets (buses) at the same time to attack the system and calculate the payoff (generation loss). Table II shows the targets and branches connected to these selected attack targets.

Table II: Target buses and branch sets. This target branches are selected from table I with maximum generation loss

Targets	Branches	Targets	Branches
z1	[10 40]	z4	[6 7 9 10 11 12 41]
z2	[1 3 5 6]	z5	[16]
z3	[16 16 17 18 19]	z6	[8 9]

Table III: Branches and generation loss due to attack in the combinations of target buses. Combinations of two targets are considered for calculation of generation loss.

	Number of branches	Generation loss (MW)
z1z2	[10 40 1 3 5 6]	51.7
z1z3	[10 40 15 16 17 18 19]	58.8
z1z4	[10 40 6 7 9 10 11 12 41]	30
z1z5	[10 40 16]	30
z1z6	[10 40 8 9]	52.8
z2z3	[1 3 5 6 15 16 17 18 19]	87.22
z2z4	[1 3 5 6 7 9 10 11 12 41]	69.59
z2z5	[1 3 5 6 16]	87.22
z2z6	[1 3 5 6 8 9]	48.22
z3z4	[15 16 17 18 19 6 7 9 10 11 12 41]	114.75
z3z5	[15 16 17 18 19]	29.70
z3z6	[15 16 17 18 19 8 9]	51.6
z4z5	[6 7 9 10 11 12 41 16]	22.25
z4z6	[6 7 9 10 11 12 41 8 9]	75.90

Table III shows the target combinations and generation losses due to the attacks. For example, from Table III, combination of target $z3$ and $z4$ combines transmission lines $15, 16, 17, 18, 19, 6, 7, 9, 10, 11, 12$ and 41. Triggering line-switching attack in these transmission lines will cause 114.75 MW of generation loss. After calculation of generation loss due to the attacks in the targets $(z1z2, z1z3, z1z4 \dots z4z6)$, defender's action is introduced. It is assumed that only one target can be protected at a time. It is also assumed that, while being protected, the branches associated with the protection set cannot be successfully attacked and will remain active. For example, if the attacker selects the combination of $z3$ and $z4$ to attack while, the defender is defending target $z4$ and $z5$, the attack will be successful for transmission lines connected to target $z3$ only. Transmission lines connected to target $z4$ will remain active as they are being defended by the defender. In this case, the generation loss will be caused by failure of transmission lines connected to target $z3$. Considering these assumptions, generation losses are calculated for all the targets against the defense sets and pay-off matrix is built.

The pay-off matrix is given in Table IV. The columns in the pay-off matrix represent the pay-offs for the attacker's strategies, and the rows are representing the strategies for the defender. Element, (i, j) in the game matrix represents the pay-off for the attack action j in response to the defense action i. From the game matrix, we can see that, if target $z1z2$ is being defended and attacked at the same time, the attack will be a failure causing no generation loss. This is why the principle diagonal of this game matrix is zero. Because all these attacks will be successfully defended by the defender. Now having a game matrix (or pay-off matrix) solution can be found in several ways (minimax theorem[22], linear programming [23] etc.).

III. Gaming : Attacker-Defender Interaction

In order to protect a set of transmission lines (i.e. any specific target/s), the defender needs to play against the attacker. In this paper, given the defender cannot defend/protect all the elements at the same time. Similarly, attacking all the elements at the same time is not possible for the attacker.

A. Solving attacker-defender two-person zero-sum game : Linear Programming

Designing an interactive decision-making game can lead to model a strategic game. For a given matrix $A_{m\times n} = \{a_{ij} : i = 1, \dots, m; j = 1, \dots, n\}$, we can consider, $\{row, i^*, column, j^*\}$ is a pair of strategies adopted by the players (attacker and defender). Then if the condition stated below is satisfied $\forall i, j$, then it can be said that the two-person zero-sum game has a saddle point in pure strategies.

$$a_{i*j} \leq a_{i*j*} \leq a_{ij*} \tag{4}$$

The strategies $\{\text{row}, i^*, \text{column}, j^*\}$ will constitute a saddle point equilibrium. They are also referred as saddle point strategies. And, the corresponding outcome of the game, $\{a_{i*j*}\}$ is termed as the saddle-point value. If a two-person zero-sum game has a single saddle point, then the value associated with that saddle point is called the value of the game. But, in case if the matrix game does not have a saddle point in pure strategies, mixed strategies are used to obtain the equilibrium solutions. A mixed strategy for a player gives a probability distribution on the space of its pure strategies. Given a $(m \times n)$ matrix

Table IV: Attacker-defender two-player zero-sum game matrix for IEEE 30 bus system. In this game matrix, pay-offs are generation losses which are calculated due to attack in the target buses. For different attack scenarios, different defending actions are also considered to calculate these pay-offs.

		Attacker														
		z1z2	z1z3	z1z4	z1z5	z1z6	z2z3	z2z4	z2z5	z2z6	z3z4	z3z5	z3z6	z4z5	z4z6	
Defender	z1z2	0	48.22	48.22	48.22	48.22	30	30	30	30	51.7	51.7	51.7	51.7	51.7	z1z2
	z1z3	29.70	0	29.70	29.70	29.70	30	58.8	58.8	58.8	30	30	30	58.8	58.8	z1z3
	z1z4	70.49	70.49	0	70.49	29.70	30	30	30	30	30	30	30	30	30	z1z4
	z1z5	22.25	22.25	22.25	0	22.25	30	30	30	30	30	30	30	30	30	z1z5
	z1z6	22.8	22.8	22.8	22.8	0	52.8	52.8	52.8	30	52.8	52.8	30	52.8	30	z1z6
	z2z3	29.70	48.22	87.22	87.22	87.22	0	29.70	29.70	29.70	48.22	48.22	48.22	87.22	87.22	z2z3
	z2z4	70.49	69.60	48.22	69.60	69.60	70.49	0	70.49	70.49	48.22	69.60	69.60	48.22	48.22	z2z4
	z2z5	22.22	87.22	87.22	48.22	87.22	22.25	22.25	0	22.25	87.22	48.22	87.22	48.22	87.22	z2z5
	z2z6	22.8	48.22	48.22	48.22	48.22	22.8	22.8	22.8	0	48.22	48.22	48.22	48.22	48.22	z3z6
	z3z4	114.75	70.49	29.70	114.75	114.45	70.49	29.70	114.75	114.75	0	70.49	70.49	29.70	29.70	z3z4
	z3z5	29.70	22.25	29.70	29.70	29.70	22.25	29.70	29.70	29.70	22.25	0	22.25	29.70	29.70	z3z5
	z3z6	51.6	.22.8	51.6	51.6	29.70	22.8	51.6	51.6	29.70	22.8	22.8	0	51.6	29.70	z4z6
	z4z5	22.25	22.25	22.25	70.49	22.25	22.25	22.25	70.49	22.25	22.25	70.49	22.25	0	22.25	z4z5
	z4z6	75.90	75.90	22.8	75.90	70.49	75.89	22.8	75.89	70.49	22.8	75.89	70.49	22.8	0	z4z6

$A = \{a_{i,j} : i = 1, \ldots, m; j = 1, \ldots, n\}$. The average value of the game can be written as:

$$J(y, w) \ = \sum_{i=1}^{m} \sum_{j=1}^{n} y_i a_{ij} w_j = y^T A w \tag{5}$$

Here y and w are the probability distribution vectors. They can be defined by:

$$\begin{aligned} y \ &= (y_1, \ldots, y_m)^T, \\ w \ &= (w_1, \ldots, w_n)^T. \end{aligned} \tag{6}$$

Here, the goal of the defender is to minimize $J(y, w)$ by an optimum choice of a probability distribution vector $y \in Y$. On the other hand, the attacker wants to maximize the same quantity by choosing an appropriate $w \in W$. The sets Y and W can be given by:

$$\begin{aligned} Y \ &= \{y \in R^m \ y \geq 0, \sum_{i=1}^{m} y_i = 1\} \\ W \ &= \{w \in R^n \ w \geq 0, \sum_{j=1}^{n} w_j = 1\} \end{aligned} \tag{7}$$

A vector w^* is a mixed strategy for the defender if the following condition is satisfied $\forall y \in Y$:

$$\overline{V}_m(A) \triangleq \max_{w \ \in W} y^{*T} A w \ \leq \max_{w \in W} y^T A w, y \in Y \tag{8}$$

Here $\overline{V}_m(A)$ is defender's average security level. In the same way, if the following inequality holds for all $w \in W$, attacker's average security level can be formulated.

$$\underline{V}_m(A) \triangleq \min_{y \ \in Y} y^T A w \ \geq \min_{y \in Y} y^T A w, w \in W \tag{9}$$

Here $\underline{V}_m(A)$ is the attacker's average security level. These two inequalities can be written alternatively as:

$$\overline{V}_m(A) \ = \min_{y} \max_{w} y^T A w, \tag{10}$$

$$\underline{V}_m(A) \ = \max_{w} \min_{y} y^T A w \tag{11}$$

For mixed strategies in two-person zero-sum game $\overline{V}_m(A) = \underline{V}_m(A)$. Thus, for a matrix game $A_{m \times n}$, equilibrium solutions can be found in mixed strategies. Average security level for both attacker and defender can be written uniquely as:

$$V_m(A) \ = \overline{V}_m(A) = \underline{V}_m(A) \tag{12}$$

For mixed strategies, to solve for equilibrium solutions, converting the game matrix into linear programming model is one of the ways. The matrix game can be expressed as : $A_{m \times n} = a_{ij}$. Here $(i = 1, 2, \ldots, m)$ and $(j = 1, 2, \ldots, n)$. All the entries in A matrix are positive $(a_{ij} > 0)$. The average game value in the mixed strategies for the attacker-defender zero-sum game can be written as:

$$V_m(A) \ = \min_{Y} \max_{W} y^T A w = \max_{W} \min_{Y} y^T A w \tag{13}$$

$V_m(A)$ is also a positive quantity which belongs to $A_{m \times n}$. This equation can be written as:

$$\min_{y \ \in Y} v_1(y) \tag{14}$$

where

$$v_1(y) \ = \max_{W} y^T A w \geq y^T A w, \forall w \in W \tag{15}$$

Additionally we can rewrite the equation as:

$$A^T y \ \leq 1_n v_1(y), 1_n \triangleq (1, \ldots, 1)^T \in R^n \tag{16}$$

Now, defender's mixed security strategy becomes,

$$\min v_1(y) \quad \text{subject to} \begin{cases} A^T \tilde{y} \leq 1_n, \\ \overline{y}^T 1_m = [v_1(y)]^{-1} \\ y = \tilde{y} v_1(y), \\ \tilde{y} \geq 0, \end{cases} \tag{17}$$

Here $\tilde{y}$ is defined as $y / v_1(y)$. This problem can be converted to a maximization problem as:

$$\max_{\tilde{y}} \tilde{y}^T 1_m \quad \text{subject to} \begin{cases} A^T \tilde{y} \leq 1_n, \\ \tilde{y} \geq 0, \end{cases} \tag{18}$$

This maximization problem will give the values of defender's mixed strategies for actions, y. This problem will take the pay-offs from Table IV as input and is subjected to the constraints from equation (18). The goal is to find the mixed strategies for the defender. This problem can be solved by using a linear programming algorithm. Meanwhile, we can write the attacker's objective function as follow,

$$\min_{\tilde{w}} \tilde{w}^T 1_n \quad \text{subject to} \begin{cases} A^T \tilde{w} \geq 1_m, \\ \tilde{w} \geq 0, \end{cases} \tag{19}$$

Here $\tilde{w}$ can be defined as $w/v_2(W)$ and v_2 can be defined as:

$$v_2 \triangleq \min_{Y} y^T A w \ \leq y^T A w, \forall y \in Y \tag{20}$$

For solving the equation (19), pay-offs from the Table IV is considered as input subject to the constraints. The outcome of this problem is attacker's mixed strategies, w associated with the game matrix from Table IV.

B. Solving attacker-defender two-person zero-sum game : Q-learning

To analyze the agent - environment interactions in Q-learning for gaming in smart grid security, the agents are the attacker and defender and the environment is the power system. The reward is the feedback from the environment for attacker-defender's action. In a two-person zero-sum game, optimal strategies can be found by the mixed strategies of all actions chosen by the participants of the game that maximize their expected long-term rewards. In this case, we consider the attacker's and defender's probability of taking an action does not change over time (stationary policy). So, we will find the convergent policies for each player at each state s. From attacker's perspective,

$$Q_A(a,d,s) \ = R_A(a,d,s) + \gamma \sum_{s' \in S} Q_A(s') T(a,d,s,s')$$

where, $Q_A(a,d,s)$ is called the quality of the state $s \in S$, $R_A(a,d,s)$ is called the reward for executing action a and d for attacker. Here, generation loss is considered as the reward, $R_A(a,d,s)$ for the attacker's and defender's action. $T(a,d,s,s')$ is the state transition probability which is considered equal for all the state transitions. The value of the game is measured by value function $V_A(s)$ which is given by:

$$V_A(s) \ = \max_{\pi_A(s)} \min_{\pi_D(s)} \sum_{a \in M_A(s)} \sum_{d \in M_D(s)} \pi_A(s) Q_A(a,d,s) \pi_D(s)$$

where, $\pi_A(s) = \pi_a(s) | a \in M_A(s), \pi_d(s) | d \in M_D(s)$. $V_A(s)$ is called the value of the state s. $Q_A(a,d,s)$ is equal to immediate reward in addition with discounted expected optimal value which can be attained from the next state s'. Here, γ is the discounted factor ranges from zero to one. It represents the impact of current decisions on long term rewards. Similarly, from the defender's perspective, defenders quality of state $Q_D(a,d,s)$ and value function $V_D(s)$ can be formulated.

In general, $V_A(s) \leq V_D(s)$ due to weak duality. Here, $V_A(s)$ and $V_D(s)$ correspond to the primal problem and dual problem, respectively. In zero-sum game, strong duality holds and we get $V_A(s) = V_D(s) = V(s)$. The game can be solved by value iteration. From the attacker's perspective, the problem becomes

$$V_A(s) \ = \max_{\pi_A(s)} \min_{d \in M_D(s)} \sum_{a \in M_A(s)} Q_A(a,d,s).\pi_a(s) \tag{21}$$

$$Q_A(s) \ = R(a,d,s) + \gamma. \sum_{s' \in S} V_A(s').T(a,d,s,s') \tag{22}$$

In this paper, it is considered that the defender's action is fixed throughout the game and initially it is determined randomly.

IV. Simulation Results and Discussions

A. Simulation Study : Linear Programming

In this section, we are going to analyze the consequences of the attack on IEEE 30 bus system. According to the assumptions made, there are 6 insecure targets ($z1, z2 \ldots z6$) and the attacker is capable of attacking two of the targets at the same time. So, the final targets are transmission lines connected to two buses at the same time. The defender is also capable of defending two targets at the same time (i.e. $z1z2, z1z3 \ldots etc.$). It is assumed that, if the attacker chooses his strategy as $\{z_i z_j\}$ (attack target i and j) and the defender chooses his strategy as $\{z_j z_k\}$ (defend target j and k), failing z_i will be successful and the generation loss will be only for switching lines connected to z_i. Payoffs in Table IV are the results of different attack and defend strategies (considering both player's action). In this section we are going to analyze the results for IEEE 30 bus system. Table IV shows that $min(max_{row}) = 29.6979$ which is not equal to $max(min_{column}) = 0$. So there is no single saddle point for solution in equilibrium and hence no values of a_{i*j*} that satisfies condition in equation 4. Having no single saddle point, the problem moves to find the proportion of times that the attacker and the defender play their own strategies. From equation 18 defender defines $\tilde{y}$, we can calculate the values of $\tilde{y}$, $y = [0\ 0.0405\ 0\ 0\ 0.4104\ 0\ 0\ 0.2076\ 0\ 0.1586\ 0\ 0.1829\ 0\ 0]$. Similarly, solving for attacker's mixed strategy we get $w = [0.0305\ 0.1988\ 0.0585\ 0\ 0\ 0\ 0.6530\ 0\ 0\ 0\ 0\ 0\ 0.0593\ 0\ 0]$.

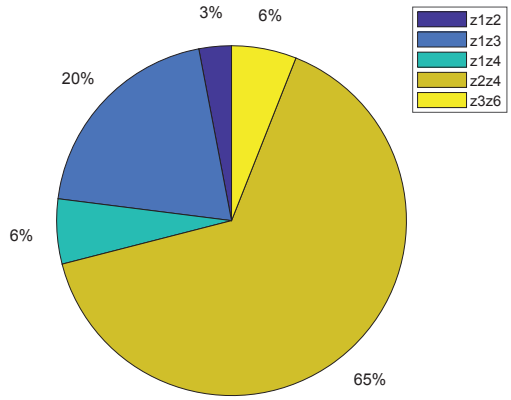

Figure 2: Attacker's mixed strategy for different attack targets in attacker-defender zero-sum two-player game for IEEE 30 bus system

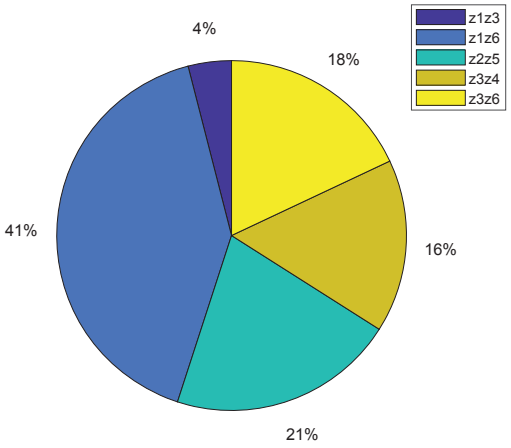

Figure 3: Defender's mixed strategy for different defending targets in attacker-defender zero-sum two-player game for IEEE 30 bus system

Figure 2 and 3 show the proportion of times that the attacker and the defender should attack and defend different targets. In figure 2, we can see that, tripping target $z2z4$ has the maximum probability of taking this action (65% probability) while target $z1z2$, $z1z3$, $z1z4$ and $z3z6$ is having the probabilities to be considered as the actions are 3%, 20%, 6% and 18% respectively. And figure 3 is showing defender's mixed strategies for the defense actions. From the figure, we can see that defending target $z1z6$ has the maximum probability of 41% to choose this target to defend. Target $z1z3$, $z2z5$, $z3z4$ and $z3z6$ are having the probabilities to be defended by the defender for 4%, 21%, 16% and 18% times of its action. For example, the attacker's strategy is $z2z4$ with 65% probability and the defender's strategy is $z1z3$ with 4% probability. With these strategies for attacker and defender obtained generation loss will be 58.8 MW. In another case, attacker's strategy of $z2z4$ (65% probability) and defender's strategy of $z1z6$ (41% probability) causes generation loss of 52.8 MW.

B. Simulation Study : Reinforcement Learning

Reinforcement learning can solve the one-shot game following the formulas from section III-B. The simulation results are given and explained here. 6 bus system has been considered as the benchmark.

Table V: Parameter information for the two-player zero-sum game between attacker and defender in 6 bus system.

Parameter	Values
Test Case	6 bus system
Number of total transmission lines	11
Number of target transmission lines	4 (30% of total transmission lines)
Maximum generation loss	210 MW
Attacker's optimal action	Transmission line - 5
Defender's fixed action	Transmission line - 2
Gamma, γ	0.9
Epsilon, ϵ	0.4
Total iterations	1000

Table V gives the value of the parameters considered for game formulation and simulation in 6 bus system. Here, epsilon, ϵ ensures that the agent in the game environment explores enough states to find the optimal action. The value of epsilon ranges from zero to one. Here, the value of $\epsilon = 0.4$ makes sure that, the agent (attacker) follows exploration for 40% of the total iterations and rest of the iterations are followed by epsilon-greedy policy. The total number of iterations considered here is 1000. This number of iterations varies according to the number of transmission lines. For smaller test power systems, the number of maximum iterations is comparatively smaller than for the bigger test power systems (such as IEEE 300 bus system). Generation loss is considered as the reward , $R(a, d, s)$ in solving two-person zero-sum game using reinforcement learning.

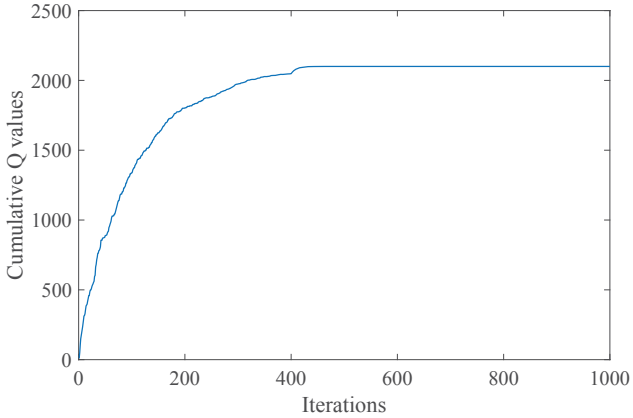

Figure 4: Attacker's cumulative Q value per iteration in the two-player zero-sum game between attacker and defender in smart grid security for 6 bus system (average of 10 runs).

Figure 4 shows the value of the game in the process of value iteration. As, the value of epsilon, $\epsilon = 0.4$, the agent will randomly explore all the possible actions within first 400

iterations to find the optimal action policy for the attacker. From the figure, the optimal action for the attacker is transmission line 5 while the defender is defending transmission line 2. This attack will cause 210 MW of generation loss. After analyzing this game result, we can conclude that, for 6 bus system, with the target of 30% of total transmission line failures, the attacker should attack line number 5 while the defender is defending line number 2. This attack will cause generation loss of 210 MW for 6 bus system. Now, we have the information that, for any randomly defended transmission line, attacker's optimal policy is attacking transmission line 5. Now we will increase the defender's strength by defending transmission line 5 and observe the attacker's optimal action. It is found that, while defending transmission line 5, the attacker chooses transmission lines 1 or 2 or 3. Because switching these transmission lines cause the same amount of generation loss. And this amount is 90.25 MW. As a result, the game value decreases and comes down to 902.5.

V. Conclusion

In this paper, we proposed two methods to solve the game in smart grid security problem. First, linear programming algorithm is used in a multi-line-switching attack scenario. Second, reinforcement learning is used for single-line-switching attack scenario. In the first case, pre-calculation of the generation loss is obtained from the system model. In the second case, we don't need any precalculation, and this solution is actually online and data-driven. Also, in reinforcement learning based solution for the one-shot game, the defender's action policy is learned from the attacker's action in the history. Linear programming shows the attacker's and defender's mixed strategy to find the optimal actions and their probability to take those actions. Reinforcement learning shows the optimal attack action in the presence of static defender's action for smart grid security.

Acknowledgment

This work is supported in part by South Dakota Board of Regents (SDBoR) Competitive Research Grant FY2018 and National Science Foundation under grant #1726964.

References

[1] S. Paul, A. Parajuli, M. R. Barzegaran, and A. Rahman, "Cyber physical renewable energy microgrid: A novel approach to make the power system reliable, resilient and secure," in *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pp. 659–664, Nov. 2016.

[2] S. Paul, M. S. Rabbani, R. K. Kundu, and S. M. R. Zaman, "A review of smart technology (smart grid) and its features," in *2014 1st International Conference on Non Conventional Energy (ICONCE 2014)*, pp. 200–203, Jan. 2014.

[3] S. Poudel, Z. Ni, T. M. Hansen, and R. Tonkoski, "Cascading failures and transient stability experiment analysis in power grid security," in *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, Sept. 2016.

[4] S. Poudel, Z. Ni, and N. Malla, "Real-time cyber physical system testbed for power system security and control," *International Journal of Electrical Power & Energy Systems*, vol. 90, pp. 124 – 133, 2017.

[5] S. Poudel, Z. Ni, X. Zhong, and H. He, "Comparative studies of power grid security with network connectivity and power flow information using unsupervised learning," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2730–2737, July 2016.

[6] S. Poudel, Z. Ni, and W. Sun, "Electrical distance approach for searching vulnerable branches during contingencies," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[7] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, pp. 86–105, Sept. 2012.

[8] P. Pourbeik, P. S. Kundur, and C. W. Taylor, "The anatomy of a power grid blackout - root causes and dynamics of recent major blackouts," *IEEE Power and Energy Magazine*, vol. 4, pp. 22–29, Sept. 2006.

[9] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, pp. 3317–3318, July 2017.

[10] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security Privacy*, vol. 9, pp. 49–51, May 2011.

[11] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[12] A. Farraj, E. Hammad, A. Al Daoud, and D. Kundur, "A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1846–1855, 2016.

[13] P. Y. Chen, S. M. Cheng, and K. C. Chen, "Smart attacks in smart grid communication networks," *IEEE Communications Magazine*, vol. 50, pp. 24–29, Aug. 2012.

[14] M. X. Cheng, M. Crow, and Q. Ye, "A game theory approach to vulnerability analysis: Integrating power flows with topological analysis," *International Journal of Electrical Power & Energy Systems*, vol. 82, pp. 29–36, 2016.

[15] Y. Xiang and L. Wang, "A game-theoretic study of load redistribution attack and defense in power systems," *Electric Power Systems Research*, vol. 151, pp. 12 – 25, 2017.

[16] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 160–169, 2013.

[17] M. J. Eppstein and P. D. H. Hines, "A random chemistry algorithm for identifying collections of multiple contingencies that initiate cascading failure," *IEEE Transactions on Power Systems*, vol. 27, pp. 1698–1705, Aug. 2012.

[18] S. Paul and Z. Ni, "Vulnerability analysis for simultaneous attack in smart grid security," in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, April 2017.

[19] J. Bialek, E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, C. Dent, I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller, M. Panteli, M. Papic, A. Pitto, J. Quiros-Tortos, and D. Wu, "Benchmarking and validation of cascading failure analysis tools," *IEEE Transactions on Power Systems*, vol. 31, pp. 4887–4900, Nov. 2016.

[20] J. Yan, H. He, X. Zhong, and Y. Tang, "Q-learning-based vulnerability analysis of smart grid against sequential topology attacks," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 200–210, Jan. 2017.

[21] Z. Ni, S. Paul, and X. Zhong, "A reinforcement learning approach for sequential decision-making process in smart grid security," in *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL'17), Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, November 2017.

[22] H. W. Corley, "Games with vector payoffs," *Journal of Optimization Theory and Applications*, vol. 47, pp. 491–498, Dec. 1985.

[23] Z. Zhu, J. Tang, S. Lambotharan, W. H. Chin, and Z. Fan, "An integer linear programming and game theory based optimization for demand-side management in smart grid," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 1205–1210, Dec. 2011.