

South Dakota State University

## Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

---

Electronic Theses and Dissertations

---

2019

### Reinforcement Learning and Game Theory for Smart Grid Security

Shuva Paul

*South Dakota State University*

Follow this and additional works at: <https://openprairie.sdstate.edu/etd>



Part of the [Power and Energy Commons](#), and the [Systems and Communications Commons](#)

---

#### Recommended Citation

Paul, Shuva, "Reinforcement Learning and Game Theory for Smart Grid Security" (2019). *Electronic Theses and Dissertations*. 3664.

<https://openprairie.sdstate.edu/etd/3664>

This Dissertation - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact [michael.biondo@sdstate.edu](mailto:michael.biondo@sdstate.edu).

REINFORCEMENT LEARNING AND GAME THEORY FOR SMART GRID  
SECURITY

BY  
SHUVA PAUL

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Electrical Engineering

South Dakota State University

2019

## MACHINE LEARNING AND GAME THEORY FOR SMART GRID SECURITY

Shuva Paul

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Electrical Engineering degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this dissertation does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Zhen Ni, Ph.D.  
Dissertation Advisor

Date

George Hamer, Ph.D.  
Head, Electrical Engineering and Computer Science

Date

Dean, Graduate School

Date

## ACKNOWLEDGEMENTS

Foremost, I would like to express the deepest gratitude to my advisor, Dr. Zhen Ni. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. Being his student is and always will be one of the greatest honors of my life. Dr. Ni gave many constructive suggestions in my study and his instructions paved the way to this dissertation. In addition, he taught me many skills in writing and revising manuscripts.

Besides my advisor, I would like to thank the rest of my dissertation committee: Dr. Steve Hietpas, Dr. Khawangee Won, Dr. Youe Zhou, and Dr. Heidi Mennenga for their insightful comments and constructive criticism at different stages of my research. I am grateful to Dr. Qiquan Qiao for practical advice and comments on the countless revision of this manuscript.

I must express my very profound gratitude to my parents and my spouse, Ananna, my sister, and sister in law, my younger brother, Hridoy, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

Finally, I am also indebted to my friends, Avijit, Rashedul, Abodh, Venkat, Priti, and many others. They have helped me stay sane through these difficult years and their support helped me overcome setbacks and stay focused on my graduate study.

# CONTENTS

LIST OF FIGURES . . . . .	xiv
LIST OF TABLES . . . . .	xix
ABSTRACT . . . . .	xx
CHAPTER 1 Introduction to cyber - physical security of smart grid . . . . .	1
1.1 Overview of the smart grid . . . . .	3
1.1.1 Traditional electric power delivery system . . . . .	3
1.1.2 Smart grid . . . . .	4
1.2 Cyber-physical security of smart grid . . . . .	7
1.2.1 Physical security of smart grid . . . . .	9
1.2.2 Cybersecurity of smart grid . . . . .	10
1.2.3 Acting towards the solution of threats . . . . .	11
1.3 Significance and organization . . . . .	13
1.3.1 Significance of the research . . . . .	13
1.3.2 Organization of the dissertation . . . . .	16
CHAPTER 2 Preliminaries . . . . .	18
2.1 Cascading failure models . . . . .	19
2.1.1 Overview . . . . .	19
2.1.2 Threat and attack model . . . . .	20
2.2 IEEE standard test systems . . . . .	22

2.3	Overview of game theory and machine learning . . . . .	25
2.4	Attack strategies . . . . .	27
2.5	Assessment metrics . . . . .	29
CHAPTER 3 Adversarial two-player zero-sum game in smart grid security . . . . .		30
3.1	Chapter overview . . . . .	31
3.2	Stage-game and multistage sequential game for power grid . . . . .	35
3.2.1	Adversarial stage-game modeling as a one-shot process . . . . .	35
3.2.2	Machine learning for multistage sequential game modeling . . . . .	48
3.3	Numerical results and simulations studies . . . . .	58
3.3.1	Stage-game in adversarial environment . . . . .	58
3.3.2	Multistage sequential game . . . . .	67
3.4	Summary . . . . .	82
CHAPTER 4 Reinforcement learning based solution for repeated game . . . . .		84
4.1	Chapter overview . . . . .	86
4.2	Problem formulation and implementation . . . . .	93
4.3	Simulation studies . . . . .	108
4.4	Summary . . . . .	137
CHAPTER 5 Deep learning for adversary game . . . . .		139
5.1	Chapter overview . . . . .	140
5.2	Proposed Framework . . . . .	142
5.2.1	Overall block diagram . . . . .	143

5.2.2	Co-simulation framework . . . . .	144
5.3	Implementation and Parameter Designs . . . . .	145
5.3.1	Deep Q-learning . . . . .	145
5.3.2	Design parameters . . . . .	150
5.4	Simulation studies . . . . .	152
5.4.1	Simulation setup . . . . .	152
5.4.2	Case studies . . . . .	153
5.5	Summary . . . . .	158
CHAPTER 6	Conclusions and opportunities . . . . .	160
LITERATURE CITED	. . . . .	162
APPENDIX	. . . . .	183

## LIST OF FIGURES

Figure 1.1.	Conceptual model of Smart Grid developed by NIST [1] . . . . .	2
Figure 1.2.	One directional flow of power in traditional grid structure . . . . .	3
Figure 1.3.	Interconnected flow of information in advanced smart grid [2] . . . . .	5
Figure 1.4.	Network architecture in the cyber-physical power system . . . . .	11
Figure 1.5.	Thesis organization tree . . . . .	16
Figure 2.1.	One line diagram of W & W 6 bus system. . . . .	22
Figure 2.2.	One-line diagram of IEEE 30 bus system. . . . .	23
Figure 2.3.	Directed graph for IEEE 30 bus system. . . . .	23
Figure 2.4.	One line diagram of New England 39 bus system. . . . .	24
Figure 2.5.	High level representation of vulnerability assessment using machine learning and game theory in power system. . . . .	26
Figure 2.6.	Simultaneous attack strategy . . . . .	28
Figure 2.7.	Sequential attack strategy utilizing Q-learning . . . . .	28
Figure 3.1.	Typical agent-environment interaction in attacker-defender two-player game. . . . .	49
Figure 3.2.	The overall diagram of attacker-defender interaction in the power system.	53
Figure 3.3.	Proposed reinforcement learning algorithm to solve the attacker-defender two-person game in power system. . . . .	54
Figure 3.4.	Transition from initial state to the new steady state through different intermediate states. . . . .	57

Figure 3.5.	(a) Attacker's and (b) defender's mixed strategies for different targets in attacker-defender zero-sum two-player game for IEEE 30 bus system	59
Figure 3.6.	Attacker's cumulative Q-value per iteration in the adversarial two-player zero-sum game for W & W 6 bus system (average of 10 runs).	61
Figure 3.7.	(a) Cumulative sum of future rewards of the attacker in the adversarial stage game (defender's protection set $\{1\}$ ) (b) probability update for all the transmission lines to be selected as an action.	62
Figure 3.8.	Action selection probabilities for the transmission lines of W & W 6 bus system.	63
Figure 3.9.	(a) Cumulative sum of future rewards and (b) action selection probabilities for the transmission lines of W & W 6 bus system with the adjusted defender's policy.	64
Figure 3.10.	(a) Action selection probabilities and (b) probability update of the attacker's action selection probabilities for IEEE 39 bus system. The randomly predefined defense action is transmission line 5.	65
Figure 3.11.	(a) Bus voltages at bus 6 and bus 31, (b) total generation and load change before and after the attack and (c) line current at transmission line 37 of IEEE 39 bus system.	66
Figure 3.12.	(a) Convergence of the attacker's number of actions and (b) generation loss due to the attacker's action in W & W 6 bus system (average of 100 runs). The defender's protection set is $\langle 1, 2 \rangle$ .	69
Figure 3.13.	Percentage of convergence to optimality for IEEE 9 bus system and IEEE 39 bus system with different exploration rates.	71

Figure 3.14. 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. (b) $\{8, 27, 37\}$ . The number of unique optimal policies here is 19. (c) Criticality index of the transmission lines of IEEE 39 bus system to be selected as the attack action by the attacker. The attack objective is to cause 60% generation loss. The defender defends transmission line $\{1, 2, 3\}$ . The number of unique optimal policies here is 35. . . . .	72
Figure 3.15. Convergence of (a) the attacker's number of actions and (b) generation loss due to the attacker's action for IEEE 39 bus system (average of 100 runs). The defender's protection set is $\langle 1, 2, 3 \rangle$ . . . . .	74
Figure 3.16. Frequency of the line attacks for IEEE 39 bus system over 100 independent runs. The defender's protection set is $\langle 1, 2, 3 \rangle$ . . . . .	76
Figure 3.17. (a) Convergence of the attacker's number of actions (average of 100 runs). The defender's protection set is $\langle 1, 2, 3 \rangle$ . (b) Frequency of the transmission line attacks for IEEE 39 bus system. The defender's protection set is $\langle 27, 29, 31 \rangle$ . (c) Convergence of generation loss due to the attacker's action (average of 100 runs). . . . .	78
Figure 3.18. Time scale for the attack simulation on IEEE 39 bus system in Power-World simulator . . . . .	80

Figure 3.19.	(a) Voltages (p.u.) at the buses connecting the target transmission lines. (b) Change in generation and load (in MW) due to the attacks in IEEE 39 bus system. . . . .	80
Figure 3.20.	(a) Line flow and (b) line current of the attacked transmission lines (8, 10, 13, and 29). . . . .	81
Figure 4.1.	Attacker-defender interaction in a repetition of a repeated game. . . . .	100
Figure 4.2.	Overall block diagram of the reinforcement learning based solution of a repeated game in smart grid security. . . . .	101
Figure 4.3.	(a) Attack flag for the repeated game. (b) Residual generation after attack-defense. . . . .	111
Figure 4.4.	(a) Players' mixed strategy payoff for individual repetitions and (b) winning percentage from the mixed strategy payoffs. The attacker's and the defender's budget is $C_B = C_D = 0.1$ % of total generation capacity. The attacking and the defending probability is $\sigma_i = \delta_i = 0.1$ . . . . .	111
Figure 4.5.	(a) The attacker's budget and (b) the defender's budget spent with each repetitions. The attacker's and the defender's budget is $C_B = C_D = 0.1$ % of total generation capacity. The attacking and the defending probability is $\sigma_i = \delta_i = 0.1$ . . . . .	112

- Figure 4.6. (a) Overall payoff of the game model and AM model with each repetitions. (b) percentage of correct, error and missing detections in the game. The attacker's and the defender's budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = \delta_i = 0.1$ . . . . . 112
- Figure 4.7. (a) Attack flag and (b) residual generation after attack-defense. The attackers' and defenders' budget  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively. . . . . 113
- Figure 4.8. (a) Players' mixed strategy payoff for individual repetitions and (b) winning percentage from the mixed strategy payoffs. The attackers' and defenders' budget was  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively. . . . . 114
- Figure 4.9. (a) Attackers' and (b) defenders' budget spent with each repetitions. The attackers' and defenders' budget was  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively. . . . . 115
- Figure 4.10. (a) Overall payoff of the game model and AM model with each repetitions and (b) percentage of correct, error and missing detections in the game. The attackers' and defenders' budget was  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively. . . . . 115

- Figure 4.11. (a) Attack flag and (b) residual generation after attack-defense. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ . . . . . 116
- Figure 4.12. (a) Players' mixed strategy payoff and (b) winning percentage from the mixed strategy payoffs. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ . . . . . 117
- Figure 4.13. (a) Attackers' and (b) defenders' budget spent with each repetitions. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ . . . . . 117
- Figure 4.14. (a) Overall payoff of the game model and AM model with each repetitions and (b) percentage of correct, error and missing detections in the game. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ . . . . . 118
- Figure 4.15. Comparison of the game model payoff and the AM model payoff for three different attack and defense strategies. . . . . 121
- Figure 4.16. Simulation loops are explained in this figure. The simulation contains repetitions, runs, and round loops. . . . . 123
- Figure 4.17. Case studies for the solution of the repeated game using reinforcement learning on IEEE 39 bus system. . . . . 124

Figure 4.18. Probability update for the optimal actions throughout the 5000 runs. . . . .	126
Figure 4.19. Probability of all the possible actions for selecting the second actions. . . . .	127
Figure 4.20. The simulation is conducted for five seconds. Starting from $t = 0$ , the first attack on line 43 is triggered at 1 seconds. Next, line 26, 20, 5, and 13 is triggered at 1.5, 2, 2.5, and 3 seconds. . . . .	135
Figure 4.21. Per unit voltages (magnitudes) at the (a) violated generators (genera- tors 2 and 3) during the attack, and (b) vulnerable generators during the attack. . . . .	136
Figure 4.22. (a) Some violated bus voltages (p.u.) and (b) generation and load loss for the whole system during the attack. . . . .	136
Figure 4.23. (a) Power flow(MW) at the target transmission lines and (b) current flow at the target transmission lines. . . . .	137
Figure 5.1. Attacker - defender two player learning and gaming interaction in power system adversarial network. . . . .	143
Figure 5.2. Co-simulation framework for learning and gaming in adversarial game in power system. The agent is learning with the help of deep learning technique executing actions in the power systems. . . . .	144
Figure 5.3. Experience replay working mechanism in deep-Q-network. DQN agent stores tuple of state, action, next state and reward in the replay buffer. From the Buffer the agent takes enough sample which are then up- dated based on randomly sampling from the batch or following greedy policy. . . . .	146

- Figure 5.4. (a) Actions to objective (number of attack actions) and (b) frequency of the transmission lines for IEEE 300 bus system. . . . . 154
- Figure 5.5. (a) Weight update from all input neuron to first hidden layer and (b) convergence of the attacking agents to optimal action strategy (1 transmission line each agent) in IEEE 39 bus system (average of 7 runs). . . 155
- Figure 5.6. (a) and (b) are showing the frequency of the appearance in the optimal target sets for the attacking agent 1 and 2, respectively. . . . . 157
- Figure 5.7. Steps to goal for multistage sequential game in IEEE 39 bus system based on (a) Q-learning and (b) deep-Q-learning. In both cases the attacker's target is to cause 5000 MW generation loss. . . . . 158

## LIST OF TABLES

Table 3.1.	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. . . .	40
Table 3.2.	Target buses and branch sets. This target branches are selected from table 3.1 with maximum generation loss . . . . .	40
Table 3.3.	Branches and generation loss due to attack in the combinations of target buses. Combinations of two targets are considered for calculation of generation loss. . . . .	41
Table 3.4.	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. . . . .	42
Table 3.5.	Parameter information for the two-player zero-sum game between attacker and defender in W & W 6 bus system. . . . .	60
Table 3.6.	Parameter information for W & W 6 bus system. . . . .	68
Table 3.7.	Number of episodes to be explored with different exploration rates for IEEE 9 bus system . . . . .	70
Table 3.8.	Number of episodes to be explored with different exploration rates for IEEE 39 bus system . . . . .	70
Table 3.9.	Parameter information for IEEE 39 bus system. . . . .	74

Table 3.10.	The attacker's action sequences for IEEE 39 bus system. . . . .	75
Table 3.11.	Comparison of cascading failures between a multistage attack and a one-shot attack. This table shows the attack actions for optimal action sequence and cascaded outages for run 1 from Table 3.10. . . . .	76
Table 3.12.	Number of converged and unique optimal strategies for the defender's different action's set for IEEE 39 bus system. . . . .	77
Table 4.1.	Parameter details for attacker-defender repeated game . . . . .	93
Table 4.2.	Different detections in the attacker-defender repeated game in smart power grid security . . . . .	98
Table 4.3.	Different RL algorithms used for different environment types in the existing multiagent RL related works. . . . .	99
Table 4.4.	Possible combinations of the players' budget and the probabilities of their actions for conducting experiments in the repeated game . . . . .	109
Table 4.5.	Observation of the game model payoff and the AM model payoff for equal budget and equal attacking and defending probabilities. The attacker's and the defender's budgets for this case studies are $0.1 \times TGC$ .	110
Table 4.6.	Observation of game model payoff and AM model payoff for equal budget and different attacking and defending probability . . . . .	113
Table 4.7.	Observation of game model payoff and AM model payoff for different attacker's and defender's budget and equal attacking and defending probability . . . . .	119

Table 4.8.	Observation of game model payoff and AM model payoff for different attacker's and defender's budget and different attacking and defending probability . . . . .	119
Table 4.9.	Observation of game model payoff and AM model payoff for different attacker's and defender's budget and equal attacking and defending probability . . . . .	120
Table 4.10.	Observation of game model payoff and AM model payoff for different attacker's and defender's budget and different attacking and defending probability . . . . .	120
Table 4.11.	Value of the parameters . . . . .	125
Table 4.12.	The attacker's and the defender's target sets containing the transmission lines from IEEE 39 branch system. Each of the target sets have 10 transmission lines. . . . .	125
Table 4.13.	Attack and defense actions with their associated probabilities, and Q-value	126
Table 4.14.	Probabilities of all the possible actions for selecting the second actions .	127
Table 4.15.	Optimal action sequences for the attacker and the defender with their probability, and Q-values. . . . .	128
Table 4.16.	Probabilities of all the possible actions for selecting the second actions .	129
Table 4.17.	Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the attacker, and the attacker and the defender has equal strength (probability). . . .	130

Table 4.18.	Probabilities of all the possible actions for selecting the second actions in favor of the attacker. The attacker and the defender both have the equal probability or strength. . . . .	130
Table 4.19.	Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the defender, and the attacker have higher strength (probability of 0.8) then the defender (probability of 0.2). . . . .	131
Table 4.20.	Probabilities of all the possible actions for selecting the second actions in favor of the defender. The attacker has higher strength (probability of 0.8) than the defender (probability of 0.2). . . . .	132
Table 4.21.	Optimal action sequences for the players with their probabilities, and Q-values. These outcomes are in favor of the defender, and the attacker have higher strength (probability of 0.8) then the defender (probability of 0.2). . . . .	133
Table 4.22.	Probabilities of all the possible actions for selecting the second actions in favor of the defender. The defender has higher strength (probability of 0.8) than the attacker (probability of 0.2). . . . .	133
Table 4.23.	Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the defender, and the attacker and the defender have equal strength (probability of 0.5).134	
Table 4.24.	Probabilities of all the possible actions for selecting the second actions in favor of the defender. Both the players have equal strength (probability of 0.5). . . . .	134

Table 5.1.	Parameter details for gaming and learning for one attacker and one de- fender . . . . .	154
Table 5.2.	Parameter details for gaming and learning for two attacker and one de- fender . . . . .	156

## ABSTRACT

## REINFORCEMENT LEARNING AND GAME THEORY FOR SMART GRID

## SECURITY

SHUVA PAUL

2019

This dissertation focuses on one of the most critical and complicated challenges facing electric power transmission and distribution systems which is their vulnerability against failure and attacks. Large scale power outages in Australia (2016), Ukraine (2015), India (2013), Nigeria (2018), and the United States (2011, 2003) have demonstrated the vulnerability of power grids to cyber and physical attacks and failures. These incidents clearly indicate the necessity of extensive research efforts to protect the power system from external intrusion and to reduce the damages from post-attack effects. *We analyze the vulnerability of smart power grids to cyber and physical attacks and failures, design different game-theoretic approaches to identify the critical components vulnerable to attack and propose their associated defense strategy, and utilizes machine learning techniques to solve the game-theoretic problems in adversarial and collaborative adversarial power grid environment.* Our contributions can be divided into three major parts:

**Vulnerability identification:** Power grid outages have disastrous impacts on almost every aspect of modern life. Despite their inevitability, the effects of failures on power grids' performance can be limited if the system operator can predict and identify the vulnerable elements of power grids. To enable these capabilities we study machine learning algorithms to identify critical power system elements adopting a cascaded failure simulator as a

threat and attack model. We use generation loss, time to reach a certain percentage of line outage/generation loss, number of line outages, etc. as evaluation metrics to evaluate the consequences of threat and attacks on the smart power grid.

**Adversarial gaming in power system:** With the advancement of the technologies, the smart attackers are deploying different techniques to supersede the existing protection scheme. In order to defend the power grid from these smart attackers, we introduce an adversarial gaming environment using machine learning techniques which is capable of replicating the complex interaction between the attacker and the power system operators. The numerical results show that a learned defender successfully narrows down the attackers' attack window and reduce damages. The results also show that considering some crucial factors, the players can independently execute actions without detailed information about each other.

**Deep learning for adversarial gaming:** The learning and gaming techniques to identify vulnerable components in the power grid become computationally expensive for large scale power systems. The power system operator needs to have the advanced skills to deal with the large dimensionality of the problem. In order to aid the power system operator in finding and analyzing vulnerability for large scale power systems, we study a deep learning technique for adversary game which is capable of dealing with high dimensional power system state space with less computational time and increased computational efficiency. Overall, the results provided in this dissertation advance power grids' resilience and security by providing a better understanding of the systems' vulnerability and by developing efficient algorithms to identify vulnerable components and appropriate defensive strategies to reduce the damages of the attack.

## CHAPTER 1 Introduction to cyber - physical security of smart grid

### ABBREVIATIONS

AMI	Advanced Metering Infrastructures.
CIP	Critical Infrastructure Protection
CVE	Common Vulnerabilities and Exposure
DOE	The U.S. Department of Energy
EISA	Energy Infrastructure Security Act
ESP	Electronic Security Parameter
GPS	Global Positioning System
IED	Intelligent Electronic Devices.
IoT	Internet of Things
ITIC	Information Technology Industry Council
MaDIoT	Manipulation of Demand over Internet of Things
NERC	North American Electric Reliability Corporation
NIST	National Institute of Standards and Technology
NSF	National Science Foundation
PMU	Phasor Measurement Units.
WAMPAC	Wide Area Monitoring, Protection, and Control

The modern era of electric power transmission and distribution system evolved from a fragile electric infrastructure improving transmission and distribution efficiency, aiding the utility operators and consumers in multidimensional ways. The advancement of this system involves, incorporation of communication devices, multidimensional communication between different sectors, cloud monitoring and control of the entire system, etc. Along with the incremental benefits and features, the vulnerability of this system is also expanding. The exposure list comprises of many critical elements forming a complex infrastructure. A large number of heterogeneous devices such as, PMU IED, and AMI are being connected to the system. Figure 1.1 shows the conceptual diagram of smart grid.

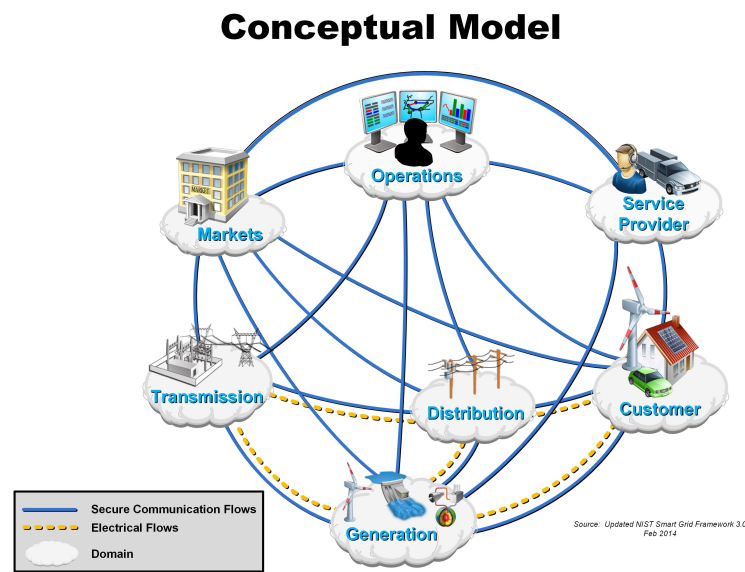


Figure 1.1. Conceptual model of Smart Grid developed by NIST [1]

The conceptual model of smart grid developed by National Institute of Standards and Technology (NIST) explains the overall composition of electric power transmission and delivery system, and its applications.

## 1.1 Overview of the smart grid

### 1.1.1 Traditional electric power delivery system

The traditional electric power delivery system is mostly electromechanical in form of infrastructure. The main problem using this electromechanical type of infrastructure is that, they cannot communicate between the devices to address any changes in the system. Moreover, in the traditional power grid, power is distributed in one way only (from the generating plant).

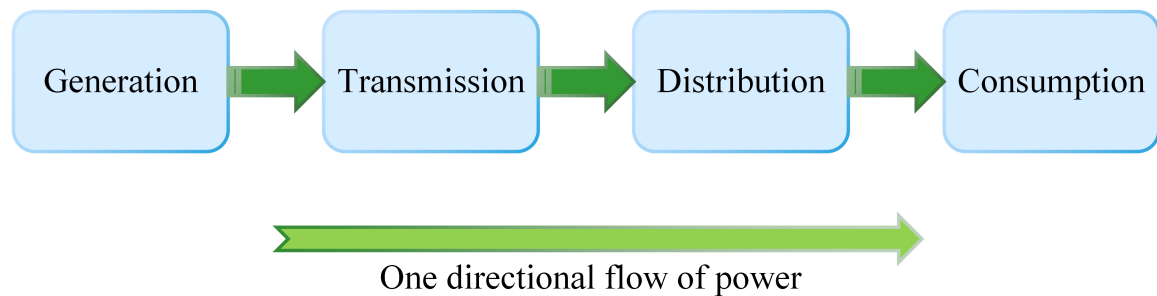


Figure 1.2. One directional flow of power in traditional grid structure

Figure 1.2, presents a simple block diagram showing unidirectional power flow in traditional electric power grid. There are four basic blocks in the power system. They are generators, transmission network, distribution network, and loads or consumers.

Traditional grid uses centralized generation which obstructs the way of using alternative energy sources. This centralized generation generates electricity from various resources. The generated electricity is delivered to the consumers via electrical transmission and distribution network. Moreover, traditional power grid uses electromechanical sensory systems which has very low data sampling speed. This obstructs the way to analyze the system for detecting any anomalies. The restoration technique after any fault or contingencies in the system is manual for the traditional power grid. Due to the aging and limitations, traditional

electric power delivery system is vulnerable to failure which can be extended to blackouts. Along with having a aging infrastructure, traditional power grid faces challenges that was never envisioned by the power system operation planner and developer before. To enhance the grid performance and resilience, the traditional fragile power grid is equipped with modern technologies which makes it a smart grid.

### 1.1.2 Smart grid

One of the major transition from traditional electric power grid to advanced smart electric power grid is the incorporation of information and communication technology (ICT) to digitize and form a network of complex systems. The transition involves solving the challenges across generation supply, transmission operation, demand variation, energy storage, renewable energies, distributed generations, market regulations, and enhancement of power system security from different internal and external threats. With the vision of modernization of electric power grid, national and international efforts are being made to increase reliability, resiliency, sustainability, and efficiency. This list also includes transition to renewable energy sources, reducing greenhouse gas emissions, implementing secure smart grid technologies and addressing cybersecurity and privacy issues. The overall objective is to develop a sustainable economy that will ensure the prosperity of future generations. Figure 1.3 shows the flow of information in the interconnected network of smart electric power grid. In the phase of modernization of electric power grid from fragile, aging, traditional form to advanced cyber-physical power system, it experienced large scale changes. Long distance transmission lines are being installed to ensure power delivery from remote generating stations to residential areas.

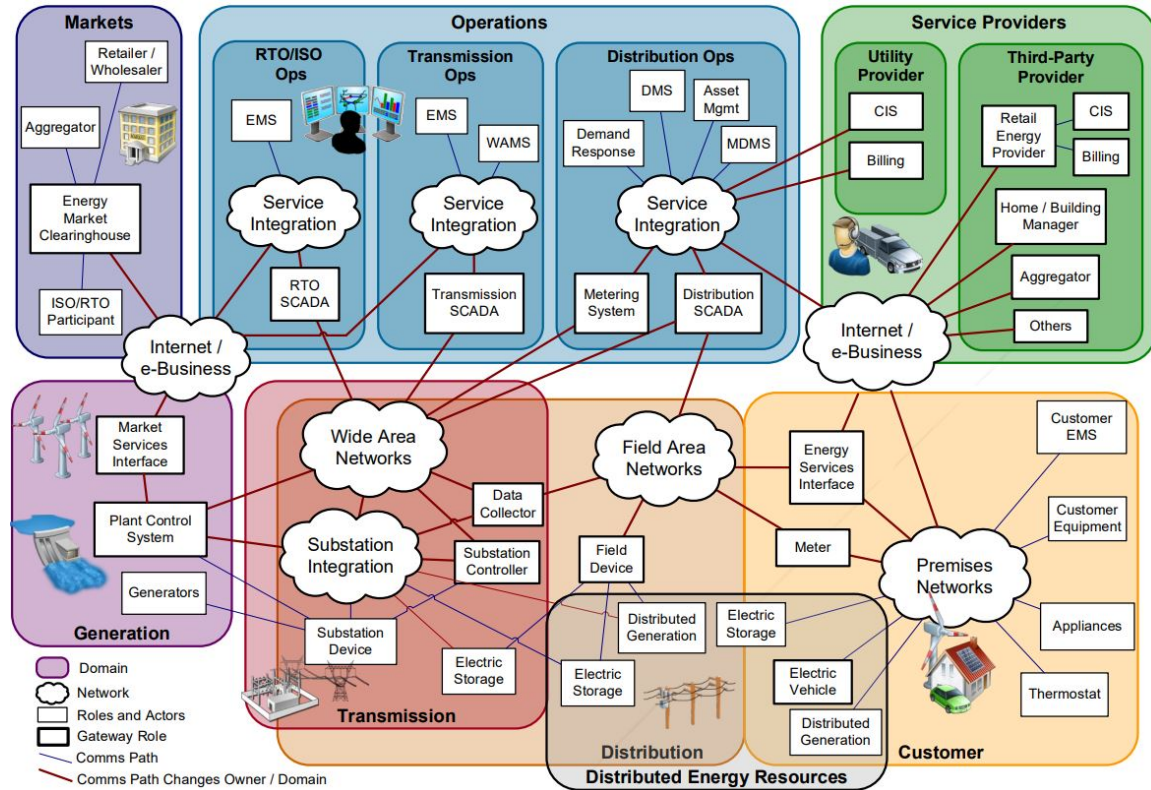


Figure 1.3. Interconnected flow of information in advanced smart grid [2]

Distributed renewable energy sources are being installed to improve environmental sustainability and economics. According to the Energy Infrastructure Security Act (EISA) of 2007 by National Institute of Standards and Technology (NIST) several benefits of smart grid has been anticipated [3]. They are, improving power reliability and quality, optimizing facility utilization and averting constriction of backup (peak load) power plants, enhancing the capacity and efficiency of existing electric power grid, improving resiliency to disruption by natural disaster and foreign attacks, enabling predictive maintenance, and “self-healing” responses to the system disturbances, etc.

Smart electric power grid with “self-healing” capability utilizes different machine learning algorithms to predict the responses of the system to different disturbances. To analyze the system more precisely and train the system to respond to any disturbances a large num-

ber of data is collected from the electric power grid. To collect these data, a large number of heterogeneous devices are being deployed. The number of data measurement devices (PMUs, AMIs, etc.) are increasing rapidly and the total number of installed Internet of Things (IoT) devices is predicted to increase from 15.41 billion in 2015 to 75.44 billion by 2025 [4]. The total number of installed PMUs in the United States had exceeded 1380 by the end of 2015. These PMUs cover almost 100% of the total US transmission system [5]. The PMUs provide high-resolution, precise, and reliable synchronous measurement data of nested power system network utilizing global positioning system (GPS). The data provided by the PMUs paves the way for wide area monitoring and protection, and control (WAMPAC) technologies. The PMU data sampling and recording rate usually varies from 10 – 60 samples per second in a 60Hz power system which is way higher than the traditional SCADA system with a sampling rate of 0.25 – 0.5 samples per second [5], [6]. This highly sampled data from the real system helps the power system operators to understand the systems dynamics more accurately. In the distribution sector, AMI is replacing once a month meter reading techniques. A large number of smart meters are being deployed to ensure multidimensional monitoring and communication.

These developments promise enhanced operational efficiency of the modern electric power generation, transmission, and delivery system. These also ensure the robustness of the grid structure. Moreover, integration and implementation of advanced technology is increasing the versatility of the electric power system to make our life comfortable.

## 1.2 Cyber-physical security of smart grid

Advanced infrastructure and multidimensional flow of information makes the smart electric power grid or cyber-physical power system vulnerable to severe security threats. Incorporating multidimensional communication, monitoring, and versatile technology, smart grid is exposed to foreign intrusion, threat to data theft, privacy violation, and major financial, social, and political damages. To make the risk management process efficient, detailed and extensive analysis of all the vulnerabilities listed in the CVE list are required [7]. The DOE's "*grid modernization*" initiatives primarily focus on the security and resiliency of the smart grid [8]. With the road-map to 2030, the DOE is promising to bring the following features in the advanced cyber-physical power system: self-healing from power disturbances, cyber-physical attack resilient operation of electric power grid, all generation and energy storage technologies utilization, ensuring consumers' active participation, etc. NSF's "*10 Big Ideas*" include different agendas involving security of the large-scale infrastructures including energy sectors [9]. CIP standards introduced by the NERC include an ESP [10]. This ESP prevents remote and unwanted intrusion to the control center or substation within its range. Despite all these standards and regulations, attacks on the power grid are increasing throughout the world. In 2018, Nigeria faced a major blackout due to the vandalizing of the major gas pipeline (70% of the electricity of this country comes from natural gas). In 2017, Nigeria faced almost 24 system collapses [11]. Colombia's national electric grid has been attacked by terrorists at a rate of 100 times each year. In addition, the United States created a list of 42 international terrorist groups as potential threats to the United State energy sector operating throughout the whole world [12]. Over the past

10 years, approximately 2,500 attacks have been conducted targeting the electric power transmission and distribution system by these terrorist groups.

In 2017, the total number of power outage related events in the united states was recorded as 3,526 and they affected almost 36.7 million people. The total duration of the power outage events in 2017 was 2,84,086 seconds which is equivalent to 197 days [12]. A recent study conducted by ITIC confirms that, just 60 minutes of downtime causes loss of \$100,000 on average. For some industries, this amount is significantly higher [13].

The motivations behind launching attack on smart electric power grid include economic benefits (e.g., electricity bill reduction), causing political unrest, create chaos and panic among the mass people, and terrorism. Along with the advanced features, the smart grid is increasing the attack surface for the threat actors. Advanced cyber-physical electric power delivery system is reaching every industrial and residential buildings and houses, letting the potential threat actors to access many of the critical elements of the grid. While incorporating heterogeneous services, cyber physical power system is becoming exposed to a wide range of threats. Having a large area of concentration, it is quite impossible to ensure security of every single subsystem.

Conducting extensive research on the security of cyber-physical power system establishes a multidisciplinary frontier, incorporating physical security of the electric power grid with cyber monitoring, control, and operation system. The outcomes of these research will help the utility operators and relevant authorities to enhance the power grid resilience, reducing the vulnerability.

### 1.2.1 Physical security of smart grid

Power outages in Nigeria, Australia, the United States, Canada, Ukraine, etc. endorse the necessity of increasing the security of electric power generation, transmission, and delivery system. Physical security of electric power grid mainly focuses on the restoration and survival of the system after experiencing contingencies. Contingency analysis is one of the core research area of power system stability to minimize the interruption of electric power delivery after experiencing contingencies. Utility operators and power system planners must go through contingency analysis to have better understanding on the events including disturbances, anomalies, faults, planned outages, etc. Contingency analysis has established security constraints for optimal power flow, unit commitment, economic dispatch, etc. These help the power system to perform a steady and efficient operation of electric power systems. For the transmission and distribution grids in the United States enforced  $N - 1$  contingency security (i.e., the power system will perform its operation security even after the loss of an components. Power grid physical security might comprises of many different and diverse factors, such as: human error, mechanical error, weather related outages, etc. In Iraq, terrorists and insurgent groups attacked power grid transmission lines to create social unrest, cause financial damages, and steal valuable transmission line materials. Columbia's electric power grid is attacked with a frequency of more than 100 times each year by the terrorists. If causing social and political unrest, panic, mass destruction, financial damage, the U.S. electric power grid would be a suitable target. So, to ensure uninterrupted power delivery to the consumers and the industries extensive research is a must for power grid communities.

### 1.2.2 Cybersecurity of smart grid

Cyber security is one of the sophisticated area of cyber-physical system. The objectives of security in smart grid include three key components, i.e., availability, integrity, and confidentiality. Figure 1.4 represents the network architecture in the cyber physical power system (inspired from [14]). Enabling multidimensional flow of information in several sectors, cyber-physical power system is serving the industry and residential community a high quality service. Moreover, high sampled measurement data allows the system operator to create situational awareness, detect anomalies and error in the system. However, the advancement in cyber-weapons are making the power system more vulnerable. Dragonfly, Dragonfly 2.0, NotPetya, WannaCry, Industroyer, and Stuxnet are among the modern generation cyber weapons [15]. Ukraine's power grid was attacked by Stuxnet malware. Ukraine experienced cyber attack twice in 2015 and 2016. The first attack affected almost 2,25,000 people. The second attack knocked down about one fifth of Kiev's total power consumption. Hackers already claimed that, they gained the access to the backdoor of some of the United State's utility companies. With these type of access they can override any command they want which can cause catastrophic damages. MadIoT is one of the recent advances in the attack strategies which manipulates the demand to cause frequency deviation, cascaded failures, and so on. MadIoT attack is initiated by botnets, such as Mirai botnet, IoTroop botnet, Ramnit botnet, etc.

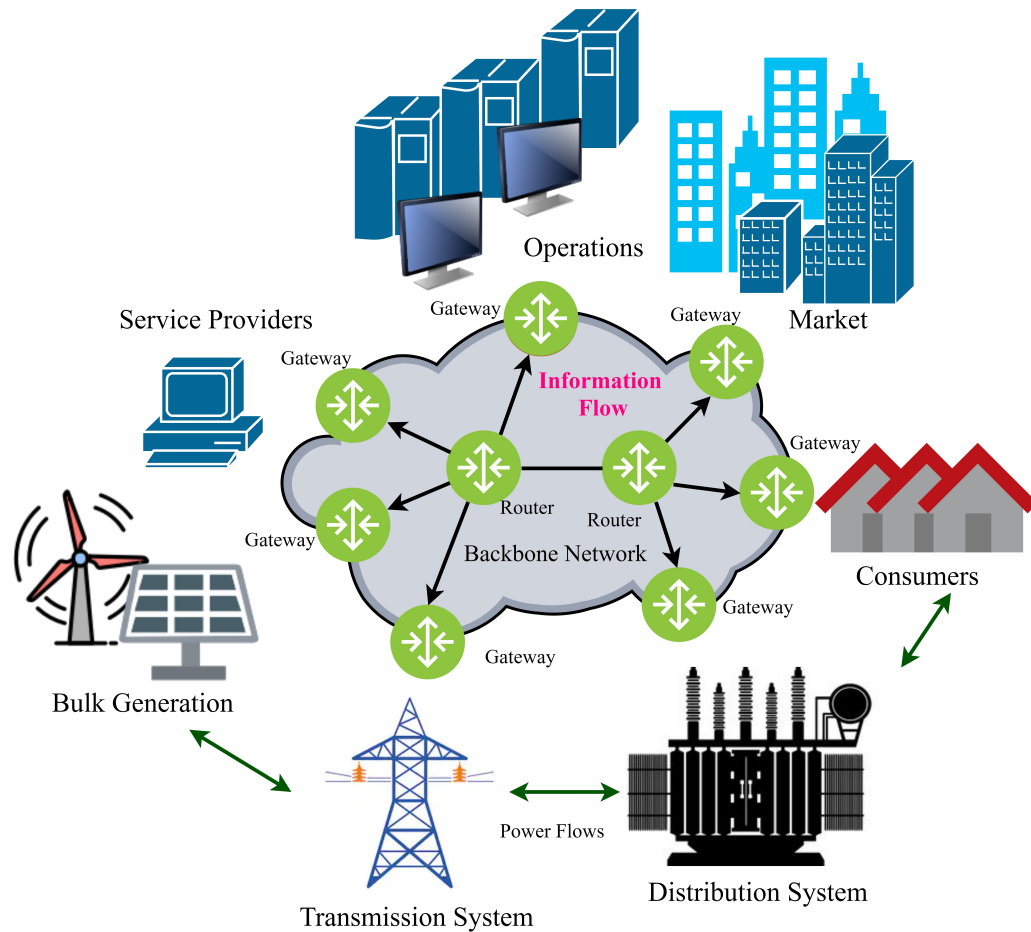


Figure 1.4. Network architecture in the cyber-physical power system

These botnets are capable of accessing the IoT devices in a network to control the smart appliances (such as, AC, water heater, dish washer, cloth dryer, etc.). WannaCry is a ransom-ware program that attacked more than 300,000 computers in almost 150 countries. According to security experts, we are heading towards *Cyber Pearl Harbor*, and the next 9/11 will be in the energy sector.

### 1.2.3 Acting towards the solution of threats

In response to the threats caused by human/non-human factors in cyber-physical power grid, numerous research attempts has been made. Machine learning and game theory is considered as the state-of-the-art techniques to address the issues in cyber-physical electric

power grid.

Machine learning methods are becoming popular in different fields of cyber-physical power system. These fields include planning, operation, maintenance, control, and security of the system [16]–[18]. Nowadays, the vulnerability of the power system has become significantly important due to the increased complexity and inter-connectivity of heterogeneous devices. Different machine learning techniques, such as different supervised, unsupervised, and semi-supervised learning algorithms are used for identify the critical contingencies of a power system. Classification and clustering techniques are used to classify and detect events, anomalies, and error data. These machine learning techniques can be integrated with game theory to solve adversarial problems in power system network.

Game theory is expected to be a formal analytical and conceptual framework with a set of mathematical tools enabling the study of complex interactions between independent rational players (i.e., the attacker and the defender). Game theory is used for explaining complex interaction and behavior of adversaries in the network (power, communication, etc.). Different games were formulated to address the challenges and issues in cyber-physical power systems (such as gas-power-water infrastructure, different communication networks [19]–[21]).

Proper solution for these game theoretic approaches are significantly important for the government authorities and the power companies to enhance the protection and avoid the consequent loss. To solve different game problems (such as general sum, zero-sum, nonzero-sum, differential etc.), different learning approaches (such as neural networks, deep neural networks, approximate dynamic programming, adaptive dynamic programming, reinforcement learning etc.) are being used by the researchers [22]–[29]. Acting

together, machine learning and game theory can respond to the adversaries in ways that reduce damage to the system or prevent the attack from happening.

### 1.3 Significance and organization

#### 1.3.1 Significance of the research

Identification of the critical contingencies and selecting the proper defense scheme against malicious attack is very crucial to enhance the security of cyber-physical power grid. Although there have been some research works exploring cyber-physical attack in the power grid, the identification of vulnerable elements in adversarial environment, proper defense scheme using machine learning approach remain a major challenge to ensure maximum security of the smart grid. To address the aforementioned issues compromising the security of smart electric power grid, this dissertation aims to assess the critical nature and vulnerability of power grid under adverse situations systematically, such as foreign intrusion, cascading failures, etc. To this extent, this dissertation focuses on identifying the critical components that leads the power system cascading failures, and in some cases massive blackouts. We utilize state-of-the-art machine learning techniques to identify the vulnerable elements considering combinations of critical elements, and sequence of critical elements. Moreover, utilizing the game theoretic approach, this dissertation proposes some novel defense strategies identifying the optimal attack and defense actions in adversarial environment.

So far, extensive research studies has been conducted investigating the security of smart grid. Cascading failures, risk graph, cyber-physical system security, and hardware/human-in-the-loop are the emergent areas or research working towards securing the modern smart

grid [30]–[35]. To elevate the investigation that makes the smart grid security stronger, different ideas have been developed including compensation theorem and current injection methods to analyze the effect of line failures, line outage distribution factors, matrix updates, contingency and influence graphs to investigate  $N - 2$  contingency problem, optimization based techniques such as mixed-integer model for  $N - k$  contingency problem, etc [36]–[45]. Identifying critical contingencies several research efforts have contributed in providing insight to develop new ideas and direction of research [46], [47]. But while analyzing vulnerability, most of the existing research fail to consider several crucial factors that could increase the risk of massive blackout, on the other hand enhance the security by identifying the vulnerable elements. To this extent, our dissertation introduces some assessment metrics to analyze the vulnerability of smart grid on the event of cascading failures, line outages, or blackouts.

While identifying the critical contingencies, the application of artificial intelligence has brought revolution in this area by increasing efficiency, accuracy, enabling near optimal prediction of events, providing defensive strategies that reduces the damage to the system from different malignant contingencies [48]–[53]. The one-shot process (either one element or combination of elements at a time) of attack cannot reveal the true nature of risk associated with the contingency analysis. Hence, this dissertation utilizes state-of-the-art machine learning techniques to identify the critical contingencies of a power system not only for one-shot process of attack, but also sequences of attack actions from the attackers' point of view.

The advanced research activities are enlightening various complex areas of power system addressing the interactions of the agents with similar/opposite goals in adversarial

environment. Game theory is playing a very significant role in these type of research where the agents execute their actions considering the power system as the game environment. A lot of research efforts has been made in power system security incorporating game theory to address power system issues [19], [54]–[62]. Cutting edge machine learning techniques are used in the game theories to address the complex interaction of agents in power system [63]–[67]. Most of the existing research works focuses on one-shot attack process and rarely consider the defender’s contribution in the process. Very few existing research works investigated the dynamic process of two players’ (i.e., the attacker and the defender) interaction in the smart grid security area, or analyzed the power system generation loss or transmission line outages. The effect of the power system operator’s defense strategies were neglected. In certain cases, the attacker has to observe the new steady-state situation and then apply a subsequent action based on the defender’s protection policy over time, and this is the dynamic game. In dynamic games, players can execute more than one action, and the time sequence has an essential role in the decision-making process [68]–[70]. On the other hand, existing game theoretic techniques for addressing the power system adversarial interactions ignores considering some crucial factors. Considering these factors might change the outcome of the game between the agents in the power system. In this extent, this dissertation explores different game theoretic approaches to incorporate different issues in adversarial gaming environment in power system. These game theoretic approaches are solved utilizing the state-of-the-art machine learning techniques which will help to reduce computational complexity and dependability, increase accuracy in determining the critical network components.

Through adequate assessment and evaluative trials, the proposed research will benefit

the advancement of monitoring, control, and operation of smart electric power grid by providing early detection, inspiring advance remedial scheme, etc. which will eventually prevent fatal damage to the grid.

### 1.3.2 Organization of the dissertation

The dissertation is organized into 6 chapters with an overall illustrated tree structure in Figure 1.5.

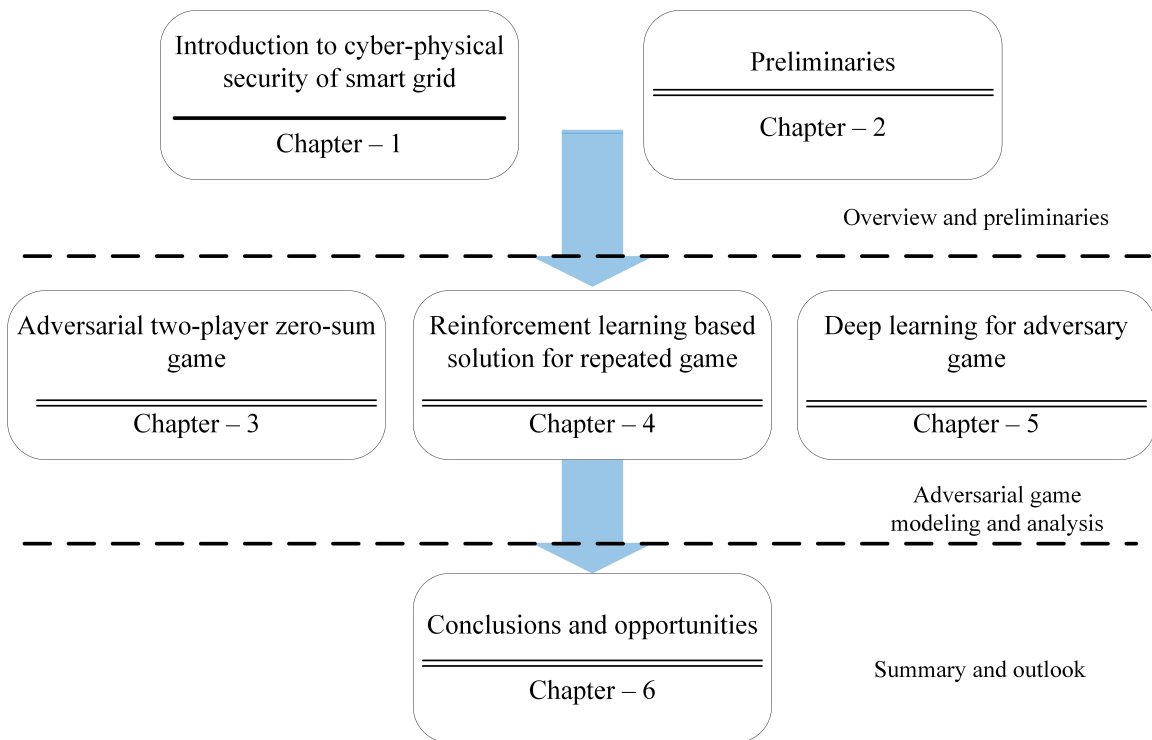


Figure 1.5. Thesis organization tree

Following Chapter 1 of introductions, Chapter 2 provides the preliminaries of the cascading failure model, assessment metrics of the vulnerability assessment, game theory, and machine learning. Chapter 3 to 5 investigates different game theoretic approaches and their solution based on machine learning techniques. The conclusions are drawn in Chapter 6 with discussions on future opportunities. Among the technical contents, Chapter 2 gives

a brief overview on cascaded failure models, discusses about the threat and attack model that this dissertation adopts. In the later part of Chapter 2, several assessment metrics are introduced which have been considered to assess the vulnerability of the system or to assess the post attack impact on the system. A brief introduction to game theory and machine learning is provided at the end of Chapter 2.

Utilizing the threat and attack model, Chapter 3, 4, and 5 introduces three different game theoretic approaches and utilizes machine learning algorithm to solve them. Chapter 3 focuses on the implementation of multistage game between the power system adversaries (the attacker and the defender). Starting with one-shot game (utilizes one-shot process of attack), this chapter implements multistage dynamic game that utilizes sequential attack process during the gaming process. Chapter 3 also compares the one-shot attack process with multistage sequential attack process. This multistage game considers passive defenders' action which adjusts its strategy from the attackers' learned action.

Addressing the limitations of multistage dynamic game, Chapter 4 implements a repeated game between the adversaries in the complex power network. Along with the simulation results, the impact of the learned attack actions is illustrated in this chapter.

In Chapter 5, a deep learning based adversary game is presented which is compatible for large scale power system to perform vulnerability assessment. Moreover, Chapter 5 shows that utilizing deep learning technique the computational efficiency can be increased.

Finally, Chapter 6 summarizes the research and discusses the future opportunities along this direction. The challenges, research methods, and technical and overall impacts throughout the process of investigation will be reviewed along with the contribution of the work.

## CHAPTER 2 Preliminaries

## NOMENCLATURE

$t$	Time steps to execute the attack.
$k$	Number of failed links.
$P_{max}$	Maximum generation capacity.
$P_{min}$	Minimum generation capacity.
$P_g$	Generated power.
$P_d$	Power demand.
$G$	Set of generator buses.
$D$	Set of load buses.
$R$	Difference between the generation and the demand.
AGC	Automatic Generation Control.
$\lambda$	Scalar factor to determine load-shedding.
$O_j$	Overload at branch $j$ .
$\bar{O}_j$	Overload threshold.
$f_j$	Power flow at line $j$ .
$\bar{f}_j$	Power flow limit at line $j$ .
CFS	Cascading Failure Simulator.
CPPS	Cyber-Physical Power System.
ML	Machine Learning.
MDP	Markov Decision Process.

## 2.1 Cascading failure models

Due to high frequency of blackout events throughout the world, cascading failure analysis in electric power grid became a very significant research topic among the researchers. A wide range of mathematical methods and analysis tools have been developed by the researcher to better illustrate the complex process of cascading failure.

### 2.1.1 Overview

Reports on historical incidents suggests that cascading failure is one of the major reasons for creating large scale blackouts [71]–[73]. Cascading failure can be defined as “the uncontrolled successive loss of system elements triggered by an incident at any location” [74]–[76]. In other words, power flows are controlled by the nature of physics and there is no limitation for capacity on the lines. But, there is a rating threshold defined for each line. If the flow is above the threshold, the transmission lines will experience thermal outage. This type of outage can alter the network configurations, giving rise to power redistribution to different path, which in turn can cause other line outage as well. If the process is repeated, it is termed as a cascading failure [77]. In [78], [79], the authors assumed that a transmission line or bus outage leads to outage of nearby buses or transmission lines with some probability. But in reality, for large scale cascading failures, an outage of a transmission line can overload remote transmission lines. The study of cascading failure modeling began with using linearized DC model and an outage rule (probabilistic) for overloaded line outages [80], [81]. The authors in [77], [82]–[86] also used similar cascade models to investigate the cascaded outage, design control strategy to mitigate the cascading failures and to identify vulnerable components in the power system [42].

The modeling of cascading failure is a very complex and challenging task. A large number of mechanisms exist that can propagate to large scale blackout just from a small set of anomalies. The models in [87], [88] presents a high level information regarding the system risk, but fails to determine network components that causes the risk. Some developed topological models for cascading outages differs from the laws of physics [89]. The other developed models, such as in [81], [90]–[94] have limitations and some cannot capture the aspects of cascading outages. They are known to be computationally intensive and expensive, and require extensive calibration. The authors in [95] proposed a cascading failure model based on dc power flow approximations which has been adopted for the researches in this dissertation as the threat and attack model.

#### 2.1.2 Threat and attack model

The cascaded failure model (threat and attack model) utilizes the line switching attack as the attack strategies. In the first step the simulator initializes the power flow. It starts the simulation with measuring the pre-contingency power flow on the selected power system test case. At  $t=1s$ , contingencies are applied. Means,  $k$  links are failed at the same time. Then, the power flow is used to calculate the system status and measure the overloads. As a result, some branches trip as a consequence of cascading failure due to overcurrent. Then re-dispatching and recalculating of power flow happens. It leads to the separation of the grid into sub-grids and re-dispatching the power flow in the sub-grids. The generators in the system are then ramped up or down to re-balance the power flow or match the supply and demand within the range of  $P_{\max}$  and  $P_{\min}$ . After re-dispatching if there is any surplus of generation in the sub grids, the system is about to trip the generators in the sub-grids

one by one. After redistributing the power flow or re-dispatching the generators. if there is more generation than load demand, the generators are tripped down to avoid surplus of generation in the islands. Usually the generators are tripped sequentially according to generation capacity, from small to large.

$$R = (\sum_{g \in G} P_g - \sum_{d \in D} P_d > 0) \quad (2.1)$$

Here 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 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 over-speed relays. Even if after this,  $R < 0$  load shedding occurs by multiplication with a scalar factor  $\lambda$ .

$$\lambda = \frac{\sum_{g \in G} P_g}{\sum_{d \in D} P_d} \quad (2.2)$$

After that, DC power flow is simulated in the simulator and the relay settings are updated. Usually time delayed over current relays are being used in the simulator to identify the branches to be tripped due to overcurrent i.e. overloads. There is a overcurrent threshold which is fixed by the system operator termed as  $\bar{o}_j$ . In the modified cascaded failure simulator 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:

$$\Delta o_j(t, \Delta t) = \begin{cases} \int_t^{t+\Delta t} (f_j(t) - \bar{f}_j) dt & \text{if } f_j(t) > \bar{f}_j \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

The simulator finds the minimum time for failing the next branch. This time is denoted

as  $\Delta T$ . Then, the time is advanced or updated with the addition of  $\Delta T$ . Then if the relay trips due to overcurrent, it will switch the online branches to offline. The overload is accumulated between time period  $t$  and  $t + \Delta t$ . The threshold overload is chosen such a way that a branch trips after a certain seconds after exceeding 50% of the branch flow limit  $\bar{f}_j$

## 2.2 IEEE standard test systems

Multi-bus power systems are considered as the benchmark systems in our research.

Figure 2.1 is showing the one line diagram of W & W 6 bus system.

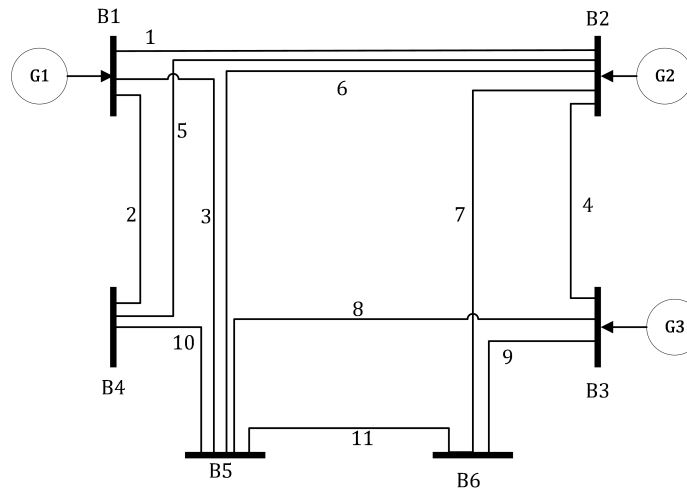


Figure 2.1. One line diagram of W & W 6 bus system.

Figure 2.2 shows the one line diagram of IEEE 30 bus system. Figure 2.3 shows the directed graph of IEEE 30 bus system. A directed graph is a graph where a set of objects (vertices or nodes) that are connected and all the edges are directed from one node to another. In Figure 2.3, nodes are representing the buses with numerical identity and the edges are representing the connection between the buses (i.e., branches or lines).

All the edges are carrying individual weights. These weights are representing the maximum power flow limit (in MW) of the branches (from buses  $\rightarrow$  buses). The connectivity (in-service or out-of-service status) can be represented by the direction of the edges and the flow limit has been represented by the weight of the edges which has been used in the simulation process of the cascading outage. Modeling of the network is the most important part in vulnerability analysis of power system. In the process of simulations, several

thousand of iterations are conducted for identification of the strongest attack combinations. To measure the system data, a cascaded failure simulator (CFS) has been adopted in some of the research works which is described in the previous section. To initiate the attack on a power system, the attacker should have topological information of a power grid. In the following researches, we will be dealing with the line switching (from in-service to out-of-service) attack. So, we have to know the status of the branches, branch connections, and flow limit for individual branches. Figure 2.4 represents the New England 39 bus system [96].

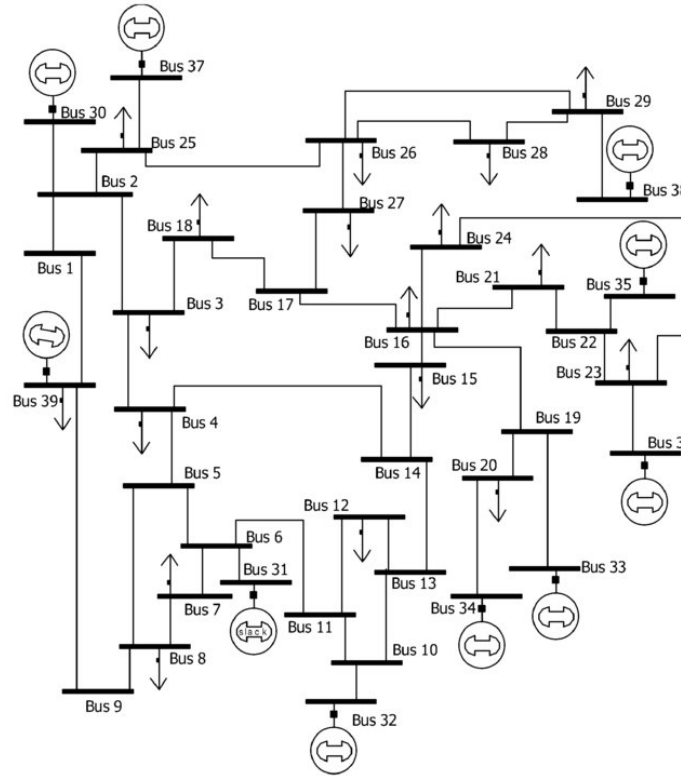


Figure 2.4. One line diagram of New England 39 bus system.

We also used IEEE 300 bus system [97] for conducting experiments in this dissertation.

### 2.3 Overview of game theory and machine learning

Game theory is expected to be a formal analytical and conceptual framework with a set of mathematical tools enabling the study of complex interactions between independent rational players (i.e., the attacker and the defender). Game theory is used for explaining complex interaction and behavior of adversaries in the network (power, communication, etc.). Different games were formulated to address the challenges and issues in cyber-physical power systems (such as gas-power-water infrastructure, different communication networks [19]–[21]). There are two major categories of games in game theory: non-cooperative game and cooperative game [98], [99]. Non-cooperative game theory can analyze the strategic decision-making process of a number of independent players, which is suitable for the smart grid attacker and defender situation. Game theory can help the power system operators to understand the attackers' motives, moves, and thus they can strengthen the security of the CPPS. A lot of research works are going on for improving the stability, security, and resiliency of smart grid under different uncertain conditions [100]–[104]. Different forms of games are formulated to solve the challenges in the CPPS. There are also two branches of non-cooperative game theory: static games and dynamic games. Static games can be seen as a one-shot process, where the players take only one action. On the other hand, dynamic games involve multiple inter dependent/independent interactions between the players. There are some other forms of games as well. Such as, repeated game, differential game, strategic honeypot game, etc. In this research, we will focus on static game, multistage sequential game, and repeated game strategy.

Machine learning methods are becoming popular along with the increasing complexity

of the critical infrastructures and their exposure to the security threats. Several machine learning (ML) algorithms are being used for detecting events and anomalies, finding critical elements of a power system, and implementing the interactions between the adversaries in a power system. These machine learning algorithms include classification and clustering methods, reinforcement learning algorithms, and artificial neural network, etc [16]–[18].

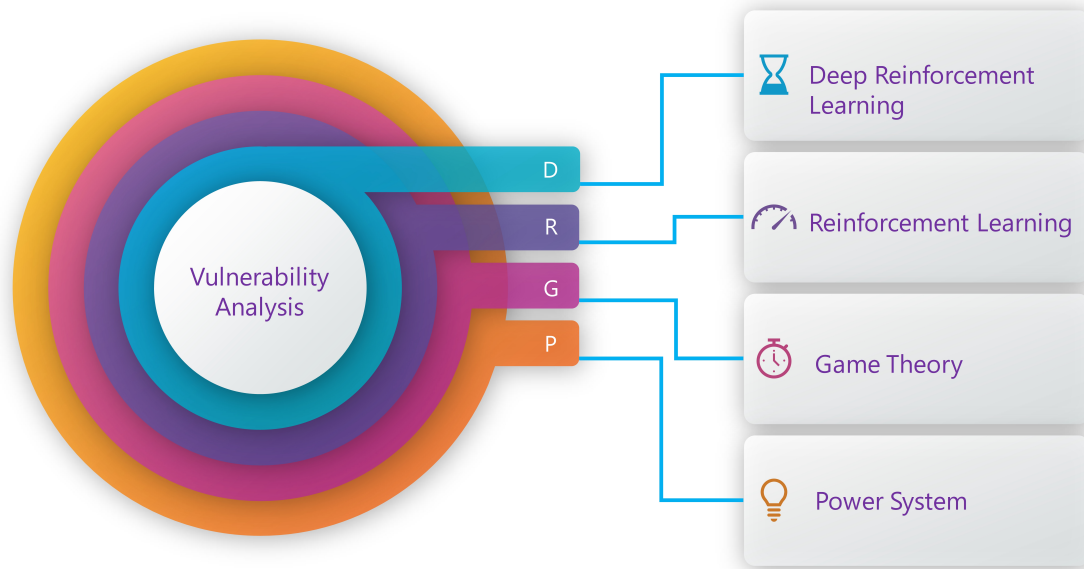


Figure 2.5. High level representation of vulnerability assessment using machine learning and game theory in power system.

Figure 2.5 shows the higher level representation of machine learning techniques and game theory to assess the vulnerability of power system. To solve different game theoretic problems, different approaches are used such as, neural networks, deep neural networks, approximate dynamic programming, adaptive dynamic programming, and reinforcement learning [22]–[29]. So, the use of game theory can aid the power system operators and planners to resolve the challenging issues of the advanced and complex electric power grid and increase the security and resiliency. Recently, several research efforts [46], [64], [105]–

[112] have been made based on Markov decision processes (MDPs), and game theory to make the security of the CPPS stronger against malicious attacks. In this dissertation, we will utilize reinforcement learning and deep-reinforcement learning techniques to solve static game, and multistage sequential game.

## 2.4 Attack strategies

Different types of attack strategies have been adopted so far to replicate the real life incidents in power system. One-shot attack [113], multi-stage attack [105], simultaneous attack [114], sequential attack [111], [115], malicious false data injection attack [116] and coordinated attack are different types of attack schemes for applying in the power grid. One-shot attack and multi-stage attack is based on the order of the attack. One-shot attack represents attacking one element at one instance or attacking a combination of elements at one instance. But multi-stage sequential attack allows the agent to execute multiple attack actions sequentially one after another. This dissertation utilizes one-shot attack, simultaneous attack and sequential attack strategy in different research problem. Figure 2.6 shows the process of simultaneous attack strategy. This attack strategy utilizes the threat and attack model which is explained and discussed in this chapter previously. The simultaneous attack strategy involves one or combination of multiple actions at a time step. This means, the attacking agent can only take action once (either one target or combination of targets). This strategy does not depend on the order of the attack.

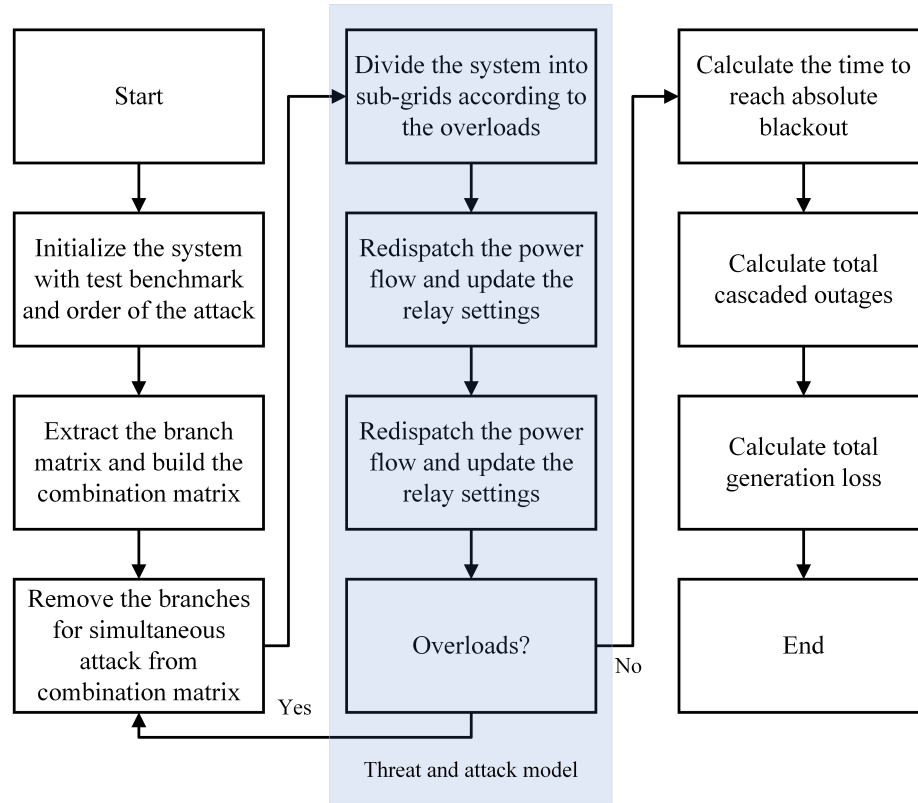


Figure 2.6. Simultaneous attack strategy

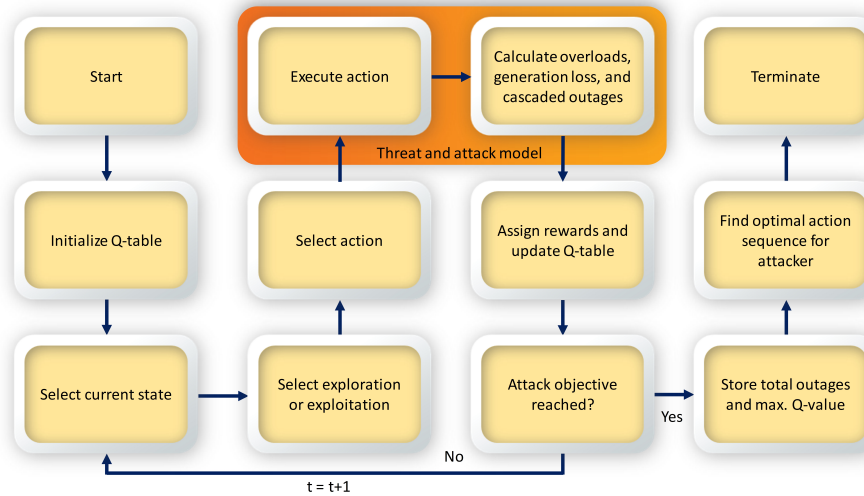


Figure 2.7. Sequential attack strategy utilizing Q-learning

Apart from the simultaneous attack strategy, we also adopt sequential attack strategy as shown in Figure 2.7. Sequential attack strategy requires multiple actions executed one

after another. In our dissertation, we adopt sequential attack strategy using Q-learning algorithm. This algorithm is a machine learning algorithm and a subset of reinforcement learning techniques. We use Q-learning algorithm with sequential attack strategy to find out the optimal sequence of actions given certain objectives.

## 2.5 Assessment metrics

To assess the vulnerability, this dissertation utilizes some assessment metrics. They are, time to reach the blackout, generation loss, total number of line outage, voltage violation etc. Time to reach absolute blackout provides the shortest time that an attacker might need to create an absolute blackout. Generation loss represents the total amount of loss that is triggered by initiating the attack. The number of total outage provides the number of total failures including the targets and the cascades. There are many more assessment metrics that can be used to assess the vulnerability of a power system. Such as, line outage distribution factor, percentage of failure, net entropy, risk, etc. These evaluation metrics can be integrated to the future version of our research.

## CHAPTER 3 Adversarial two-player zero-sum game in smart grid security

## NOMENCLATURE

$z$	Targets for the attacker
$A$	Game matrix
$J$	Average value of the game
$y$	Defender's probability distribution vector
$w$	Attacker's probability distribution vector
$t$	Time step
$a$	Attack action
$d$	Defense action
$A$	Attack action set
$D$	Defense action set
$\gamma$	Discount factor
$s_t(l)$	Status of line $l$ at time $t$
$N_0$	Number of transmission line outage
$N_\theta$	Attack objective
$k$	Number of actions executed
$\varepsilon$	Exploration probability
$V_m$	Average game value in the mixed strategies
$Q_A/Q_{att}$	Attacker's quality of state S due to attack and defense actions
$Q_D/Q_{def}$	Defender's quality of state S due to attack and defense actions
$\pi_A$	Attackers' attack policy/strategy
$\pi_D$	Defenders' defense policy/strategy
$V_A/V_{att}$	Attackers' value of state S
$V_D/V_{def}$	Defenders' value of state S
$T(a, d, s, s')$	Transition probability
$Pr(a, d, s)$	Probability of a state-action pair to be selected
$R^A$	Reward assigned for executed attack action
$R^D$	Reward assigned for the executed defense action

### 3.1 Chapter overview

The modern electrical power grid is highly interconnected and coupled. Large-scale power system outage is likely to happen due to the consequence of cascading failures, which usually start with a single transmission line outage [117]–[119]. There are many research projects investigating the vulnerability of different components in the power system from the viewpoints of communication, information security, stability, control and many more. The heterogeneous nature of smart grid inspires the integration of advanced techniques to overcome various technical challenges at different levels, such as design, control, and security [120]–[123]. Cascading failures, risk graph, cyber-physical system security, and hardware/human-in-the-loop are the emergent areas or research working towards securing the modern smart grid [30]–[35]. Apart from the attackers in the smart electric power grid, power system operators, different protection schemes acts as the defender of the system. Game theory is used for the implementation of complex interactions between the adversaries (the attacker and the defender) in a smart electric power grid. In Chapter 2, we have introduced different evaluation metrics, simultaneous and sequential attack schemes, and Q-learning for identifying the sequence of critical contingencies from a power system model from the perspective of an attacker.

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 [68]. In [59], 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 [124]. In [125] and [126], 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 [62], power system measurements are attacked in a static game to investigate the interactions between attacker and defender. Apart from the aforementioned categories of game theory, there are cooperative and non-cooperative game theories [98], [99]. Non-cooperative game theory can analyze the strategic decision-making process of a number of independent players, which is suitable for the smart grid attacker and defender situation. In the existing literature most of the existing works are based on one-shot game where no dynamic interactions were considered between the players. Moreover, the post attack impacts in the power system have been rarely reported in the existing literature. Very few existing papers investigated the dynamic process of two players' (i.e., the attacker and the defender) interaction in the smart grid security area, or analyzed the power system generation loss or transmission line outages. The effect of the power system operator's defense strategies were neglected. In certain cases, the attacker has to observe the new steady-state situation and then apply a subsequent action based on the defender's protection policy over time, and this is the dynamic game. In dynamic games, players can take more than one action, and the time sequence has an essential role in the decision-making process [68]–[70]. In this multi-stage game, choosing a proper action is quite a challenging task. In order to reach the optimal action policy, the players must follow well-formulated rules based on the full observation of the system. For example, the adaptive dynamic programming/reinforcement learning algorithm was used in the control system to find online equilibrium solutions for a two-player zero-sum game

over time [127]–[130]. Both dynamic programming and reinforcement learning basically solve the same Bellman’s equation. In dynamic programming, it is assumed that the model is known. Dynamic programming requires the knowledge of Markov decision process or a model of the environment so that the recursions can be carried out. It typically falls under “planning” rather than “learning”. In Q-learning, it accesses a set of samples. These samples can be collected online or offline. Q-learning is basically a relaxation process which goes from experience to a policy.

Inspired by the gaming in smart grid security researches, we propose solutions for the one-shot game and multistage game in smart electric power grid. Based on the aforementioned discussion, we propose a two-person zero-sum game model in smart grid security. Two-person zero-sum game can possibly be solved using mini-max or maxi-min theorem, Nash equilibrium, reinforcement learning, neural networks etc. We adopt the Q-learning approach to solve the two-person zero-sum game because Q-learning supports dynamic online learning. This solution will open a new direction to study smart grid security applications between the attacker and the defender. It will also generate multi-disciplinary research among several societies. The contributions of this chapter are summarized below:

1. First we propose two solutions for the one-shot game/stage game in smart grid adversarial environment. Linear programming based solution provides both the attacker’s and the defender’s optimal action strategies. Reinforcement learning based solution provides optimal action strategy and alternative action choices for the players (in case of limited availability of the actions). Finally we reported the attack impacts on the power system utilizing the Power World simulator.

2. Later, we implement and propose a new solution for the multistage sequential game in smart grid security based on reinforcement learning. Instead of a one-shot game where the attacker and the defender will choose their actions based on the static payoff matrix, we are adopting a learning scheme to find out the optimal attack sequences and identify the most vulnerable transmission lines. Our proposed solution considers the defender's action and finds the attacker's optimal action sequences (critical transmission lines of power system) playing against the defender.
3. The learned optimal attack actions will then be used as a protection set for the defender. The new multistage game will be conducted, and the attacker will have fewer choices to achieve the objectives. The new multistage game is compared to the previous game and the results indicate that the proposed learning algorithm could enhance the power system protection by adopting the attacker's learned policy. Thus, the power system operator can make timely protection policy adjustments to avoid the "smart" attack.
4. The multistage attack is compared with the one-shot attack. A multistage attack creates more transmission lines outage than a one-shot attack. The proposed reinforcement learning solution for smart grid security can also use the generation loss as the reward feedback in addition to the line outages as described above.
5. We explore the effects of learning parameter on the convergence to the optimal strategies for the proposed adversarial multistage sequential game. We show that increasing exploration rate increases the percentage of converged optimal strategies that solve the game. We also show step by step how adjusting the defender's defense

policy limits the attacker's available attack window. Then we illustrate the impact of the learned attack policies in terms of transmission line outages, bus voltage violations, and generation loss on a simulated power system platform using PowerWorld simulator.

The outcomes of this chapter is expected to aid the power system operators and planners to improve the defense strategy and minimize loss for both one-shot attack and multistage attack in power system. In this chapter we will introduce the defender as a passive learner. We consider both one-shot game and multistage sequential dynamic game for mimicking the interactions between the adversaries in the power system. In 3.2.1 we implement one-shot game between the adversaries and solve the game using linear programming and reinforcement learning. In 3.2.2, we implement multi-stage sequential dynamic game and provide reinforcement learning based solution.

## 3.2 Stage-game and multistage sequential game for power grid

### 3.2.1 Adversarial stage-game modeling as a one-shot process

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. Both single 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 adjusted from attacker's previous action history. This adjustment will strengthen the security of power system by identifying and protecting critical elements, reducing generation and financial loss.

### 3.2.1.1 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 represented 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 [95] [114]. Generation loss is also used as the reward for reinforcement learning based problem formulation and solution.

**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. W & W 6 bus system is used to demonstrate the proposed solution of the game using reinforcement learning. Transmission line indices are used to represent the attacks in the power system test cases.

**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 solutions. To calculate the generation loss, threat and attack model is adopted [74], [95], [114]. In this chapter, the threat and attack model 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 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 3 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 threat and attack model described in Chapter 2. Then the model executes the actions (switches the transmission lines from active status to inactive status) and calculates the generation loss, bus separation, transmission line outages (cascading failures). Thus, the generation loss is calculated. These generation losses are used both in the solution of the game using linear programming and reinforcement learning.

---

**Algorithm 1: Calculation of generation loss matrix**


---

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

```

1 Initialization;
2 for Given test case do
3   Load the test case;
4   Extract the attack matrix;
5   for Given attack matrix do
6     Run power flow;
7     Switch the branches from attack matrix;
8     Divide into sub-grids according to the overloads;
9     Redispatching the power flow;
10    Update the relay settings;
11    if There are overloads then
12      Trip the branches according to updated settings;
13      Check for the overloads again;
14    else
15      Calculate total generation loss;
16    end
17  end
18  Display the generation loss matrix;
19 end
  
```

---

**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 3.1 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 3.1, 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 3.1. 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 (MW)	Bus	Branches	Loss (MW)
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 3.1, we select buses 8, 7, 2, 12, 6 and 13 as targets for attacking and defending. We name them as  $z_1, z_6, z_2, z_3, z_4$  and  $z_5$  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 3.2 shows the targets and branches connected to these selected attack targets.

Table 3.2. Target buses and branch sets. This target branches are selected from table 3.1 with maximum generation loss

Targets	Branches	Targets	Branches
$z_1$	[10 40]	$z_4$	[6 7 9 10 11 12 41]
$z_2$	[1 3 5 6]	$z_5$	[16]
$z_3$	[16 16 17 18 19]	$z_6$	[8 9]

Table 3.3. 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 3.3 shows the target combinations and generation losses due to the attacks. For example, from Table 3.3, 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...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.

Table 3.4. 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	
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	z2z6
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	z3z6
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

The pay-off matrix is given in Table 3.4. 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 [131], linear programming [132] etc.).

### 3.2.1.2 Attacker-defender interaction in gaming

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.

#### **Solving attacker-defender two-person zero-sum game using 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.

$$a_{i^*j} \leq a_{i^*j^*} \leq a_{ij^*} \quad (3.1)$$

The strategies  $\{row, i^*, column, j^*\}$  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  $A = \{a_{i,j} : i = 1, \dots, m; j = 1, \dots, n\}$ . The average value of the game can be written as:

$$J(y, w) = \sum_{i=1}^m \sum_{j=1}^n y_i a_{ij} w_j = y^T A w \quad (3.2)$$

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

$$\begin{aligned} y &= (y_1, \dots, y_m)^T, \\ w &= (w_1, \dots, w_n)^T. \end{aligned} \quad (3.3)$$

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:

$$\begin{aligned} Y &= \{y \in R^m \mid y \geq 0, \sum_{i=1}^m y_i = 1\} \\ W &= \{w \in R^n \mid w \geq 0, \sum_{j=1}^n w_j = 1\} \end{aligned} \quad (3.4)$$

A vector  $w^*$  is a mixed strategy for the defender if the following condition is satisfied

$\forall y \in Y$ :

$$\bar{V}_m(A) \triangleq \max_{w \in W} y^{*T} A w \leq \max_{w \in W} y^T A w, y \in Y \quad (3.5)$$

Here  $\bar{V}_m(A)$  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.

$$\underline{V}_m(A) \triangleq \min_{y \in Y} y^T A w \geq \min_{y \in Y} y^T A w, w \in W \quad (3.6)$$

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

$$\bar{V}_m(A) = \min_y \max_w y^T A w, \quad (3.7)$$

$$\underline{V}_m(A) = \max_w \min_y y^T A w \quad (3.8)$$

For mixed strategies in two-person zero-sum game  $\bar{V}_m(A) = \underline{V}_m(A)$ . 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:

$$V_m(A) = \bar{V}_m(A) = \underline{V}_m(A) \quad (3.9)$$

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  $(i = 1, 2, \dots, m)$  and  $(j = 1, 2, \dots, n)$ . 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:

$$V_m(A) = \min_Y \max_W y^T A w = \max_W \min_Y y^T A w \quad (3.10)$$

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

$$\min_{y \in Y} v_1(y) \quad (3.11)$$

where

$$v_1(y) = \max_W y^T A w \geq y^T A w, \forall w \in W \quad (3.12)$$

Additionally we can rewrite the equation as:

$$A^T y \leq 1_n v_1(y), 1_n \triangleq (1, \dots, 1)^T \in R^n \quad (3.13)$$

Now, defender's mixed security strategy becomes,

$$\begin{array}{ll} \min v_1(y) & \\ \text{subject to} & \left\{ \begin{array}{l} A^T \tilde{y} \leq 1_n, \\ \tilde{y}^T 1_m = [v_1(y)]^{-1} \\ y = \tilde{y} v_1(y), \\ \tilde{y} \geq 0, \end{array} \right. \end{array} \quad (3.14)$$

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

$$\begin{array}{ll} \max_{\tilde{y}} \tilde{y}^T 1_m & \\ \text{subject to} & \left\{ \begin{array}{l} A^T \tilde{y} \leq 1_n, \\ \tilde{y} \geq 0, \end{array} \right. \end{array} \quad (3.15)$$

This maximization problem will give the values of defender's mixed strategies for actions,  $y$ . This problem will take the pay-offs from Table 3.4 as input and is subjected to the constraints from equation (3.15). 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,

$$\begin{aligned} \min_{\tilde{w}} \quad & \tilde{w}^T 1_n \\ \text{subject to} \quad & \begin{cases} A^T \tilde{w} \geq 1_m, \\ \tilde{w} \geq 0, \end{cases} \end{aligned} \quad (3.16)$$

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

$$v_2 \triangleq \min_Y y^T A w \leq y^T A w, \forall y \in Y \quad (3.17)$$

For solving the equation (3.16), pay-offs from the Table 3.4 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 3.4.

**Solving attacker-defender two-person zero-sum game using 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,  $\pi_A(s) = \pi_a(s) | a \in M_A(s)$ ,  $\pi_D(s) = \pi_d(s) | d \in M_D(s)$ .  $V_A(s)$  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,  $\gamma$  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

$$V_A(s) = \max_{\pi_A(s)} \min_{d \in M_D(s)} \sum_{a \in M_A(s)} Q_A(a, d, s) \cdot \pi_a(s) \quad (3.18)$$

$$Q_A(s) = R(a, d, s) + \gamma \cdot \sum_{s' \in S} V_A(s') \cdot T(a, d, s, s') \quad (3.19)$$

In this paper, it is considered that the defender's action is fixed throughout the game and initially it is determined randomly. This problem can be solved by using value iteration

[126], [133].

$$\begin{aligned}
& \max_{\pi} V_a(s') \\
& s.t. \sum_{a \in S_{N_a}} Q(s, a, d) \pi \geq V_a(s') \\
& \sum_{a \in S_{N_a}} \pi(a) = 1 \\
& \pi(a) \geq 0, \forall a \in S_A
\end{aligned} \tag{3.20}$$

The optimal policy can be defined as

$$\pi'(s) = \arg \max_a Q_{\pi}(s, a, d) \tag{3.21}$$

We update the probabilities of the state-action pairs following the formula below

$$Pr(s, a, d) \leftarrow \frac{C(s, a, d)}{\sum_{a \in A, d \in D} C(s, a, d)} \tag{3.22}$$

where,  $C(s, a, d)$  represents the number of times state  $s$  is visited by the agent (the attacker) while taking action  $a \in A$ . This probability is calculated based on the frequency of that specific state action pairs visited. The attacker's mixed strategy for a given state  $s$  will be:

$$\pi_A(s) = [Pr\{a(s) = a_1\}, \dots, Pr\{a(s) = a_N\}] \tag{3.23}$$

where

$$\sum_{i=1}^N Pr\{a(s) = a_i\} = 1 \tag{3.24}$$

Here  $Pr\{a(s) = a_i\}$  is the probability of choosing attack action  $a_i$  in state  $s \in S_A$ .  $\pi_A(s)$  is the probability distribution over the attacker's action space associated with state  $s$ .

### 3.2.2 Machine learning for multistage sequential game modeling

In multistage sequential game, the players are allowed to take limited number of actions in the environment. Multistage sequential game based on Q-learning utilizes Q-learning algorithm (a reinforcement learning algorithm) to identify the optimal attack actions from the attacker's perspective.

### 3.2.2.1 Problem formulation

In this section, we discuss the necessary definitions used to conduct the game. Then we analyze the gaming problem formulation between the adversaries. Two-player games are natural extensions from single agent Markov decision process to multiagent systems. In this subsection, we will provide the definition and the formulation of a two-player game in the smart grid.

**Definition 2:** A 2-player Markov game has a tuple defined as  $\langle S, A, D, R^A, R^D \rangle$ , where  $S$  is a discrete state space,  $A$  and  $D$  are the discrete action spaces of players 1 and 2.  $S \times A \times D \rightarrow R$  is the payoff function for each player [68], [134].

There are two players who make the actions in turn: player 1 (the attacker) and player 2 (the defender). At state  $s$ , the players choose their actions  $a \in A, d \in D$  independently and receive rewards  $R^A(s, a, d)$  and  $R^D(s, a, d)$ , respectively. If

$$R^A(s, a, d) + R^D(s, a, d) = 0 \quad (3.25)$$

the game is called a zero-sum game. If condition in equation (3.25) is not satisfied, then the game is termed as a *general-sum* game.

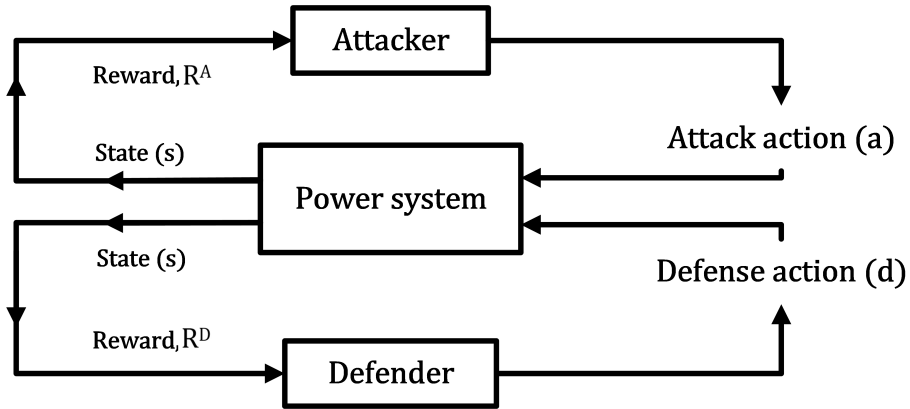


Figure 3.1. Typical agent-environment interaction in attacker-defender two-player game.

Figure 3.1 shows a modified agent (i.e., the attacker and the defender) - environment (power system) interaction for a two-player game framework. The attacker obtains system state  $s$  and applies the attack action  $a$ , and it will receive reward  $R^A$ . Meanwhile, the defender will conduct the same process and receive reward  $R^D$ . The objective of the attacker (player 1) is to maximize the discounted sum of future rewards, while the objective of the defender (player 2) is to minimize it.  $\pi_A$  and  $\pi_D$  are the strategies of attacker and defender, respectively.

For an initial state  $s$ , the attacker and the defender acquire the game values as

$$V_{att}(s, \pi_A, \pi_D) = \sum_{t=0}^{\infty} \gamma^t E(R_t^A \mid \pi_A, \pi_D, s_0 = s) \quad (3.26)$$

$$V_{def}(s, \pi_A, \pi_D) = \sum_{t=0}^{\infty} \gamma^t E(R_t^D \mid \pi_A, \pi_D, s_0 = s) \quad (3.27)$$

where,

$$\pi_A(s) = \{\pi_a(s) \mid a \in A\}$$

$$\pi_D(s) = \{\pi_d(s) \mid d \in D\}$$

$V_{att}$  and  $V_{def}$  represent the game value for the attacker and the defender, respectively.  $R_t^A$  and  $R_t^D$  represent the reward assigned for the attacker and the defender, respectively.  $\pi_A^*$  and  $\pi_D^*$  are the optimal strategies for the attacker and the defender, respectively. Relatively smaller value of  $\gamma$  (greater than zero) indicates higher weights of immediate rewards, whereas larger value (closer to one) of  $\gamma$  emphasizes on the future rewards. With the multi-stage game, the attacker has an option to choose different attack targets in a time sequence. This gives the attacker flexibility to choose different attack sequences. The defender is the power system operator or the security administrator who monitors the power system's cyber and physical networks.

A Nash equilibrium for two-player Markov game can be defined as follows.

**Definition 3:** In a two-player Markov game, a Nash equilibrium point is a pair of strategies  $(\pi_A^*, \pi_D^*)$  such that for all  $s \in S$  [134]

$$V_{att}(s, \pi_A^*, \pi_D^*) \geq V_{att}(s, \pi_A, \pi_D^*), \forall \pi_A \in \Pi_A \quad (3.28)$$

and

$$V_{def}(s, \pi_A^*, \pi_D^*) \geq V_{def}(s, \pi_A^*, \pi_D), \forall \pi_D \in \Pi_D \quad (3.29)$$

Here, both of the players have complete information about each others payoff function. In Nash equilibrium, each agent's strategy has to be the best response to the opponent's strategy. The definition of Nash equilibrium is similar in other types of games. The two players repeatedly play games according to their knowledge until the end. In this game, if any (active) player reaches the Nash equilibrium point, it also means that this player obtains the optimal policy.

To solve this game and find the optimal strategies for two players, one widely used solution principle is the closed-loop Nash equilibrium [135], [136]. A Nash equilibrium can be defined as a state of the game where no player can change its reward by deviating unilaterally from the equilibrium state. For a static game, the existence of the Nash equilibrium is guaranteed by Nash's theorem [135]. In a multistage dynamic game, the existence of the Nash equilibrium can be obtained through the decision-making process over time. Q-learning is one of the ideal ways to find the players' optimal policies (Nash equilibrium). Note, there could be multiple optimal policies (Nash equilibrium) for dynamic games [133], [137]. A Nash equilibrium is a concept of a solution that gives a steady state condition of that specific game. None of the players would prefer to change their strategies as that would minimize the payoffs [138]. Also, in [127], a two-player zero-sum game

in which only one of the players is learning and optimizing its strategy while the opponent player is maintaining its fixed policies. This ended the learning procedure with the calculation of desired Nash equilibrium solution of the game. Given that, the defender in our two-player zero-sum game is passive. So while the attacker converges to its optimal action strategy facing the passive defender and maximizes its payoff, none of the players can deviate from their action strategies as deviation will minimize their own payoffs. This confirms the existence of Nash equilibrium in the proposed game.

### 3.2.2.2 Proposed reinforcement learning solution

In this section, we propose a new solution for the two-player zero-sum dynamic game based on reinforcement learning between the attacker and the defender in the applications of the smart grid. First, we introduce the overall framework and algorithm flowchart, the design of reward feedback and  $Q$ -function and then provide the convergence analysis.

**Overall diagram and algorithm flowchart** The overall diagram of attacker-defender interaction in a power system appears in Figure 3.2. The game starts with the normal operating condition (i.e., steady-state) of the power system. Assuming that the defender uses a passive defense scheme, and the attacker uses an active learning attack scheme, the attacker will achieve its attack objective by learning and creating certain attack action sequences through the trial-and-error process. The attacker does not have a first action advantage in this dynamic game. The defender's policy is pre-defined at the beginning of individual games. Thus, the attacker's action is taken against the defender's action. At the end of the multistage game, the defender observes total transmission line outages, and generation losses and adjusts its action strategy from the attacker's learned action

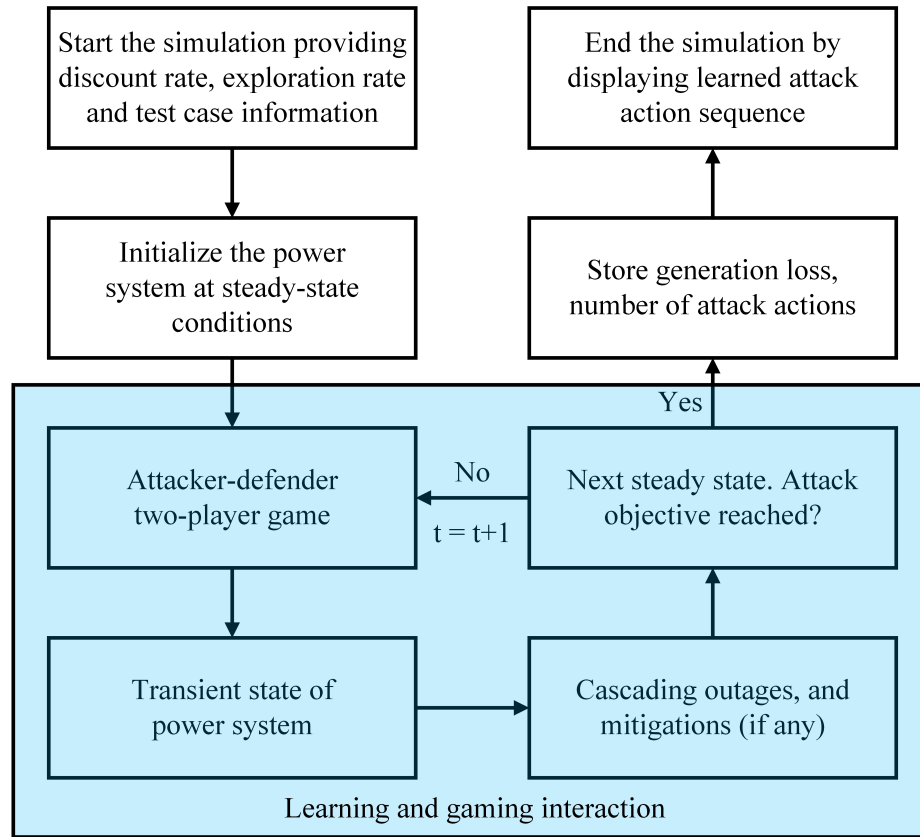


Figure 3.2. The overall diagram of attacker-defender interaction in the power system.

sequence. During the post-attack stage, the system may have cascading failures. A typical power system operator tries to mitigate the cascading failure and moves the system to a new steady-state condition by generator re-dispatching, islanding, etc. An attacker cannot attack (trip) the protected transmission lines directly. However, the protected transmission lines may be automatically tripped due to cascading failure. Figure 3.3 shows our proposed reinforcement learning algorithm to solve the zero-sum game between the attacker and the defender. The game starts by initializing the  $Q$ -table and the defender's set of actions. At the beginning, the current state is initialized as the normal operating condition of the given power system. At this state, all the transmission lines are active. A random number is generated to select the probability of exploration. The exploration probability of the attack

action is denoted by epsilon ( $\epsilon$ ).

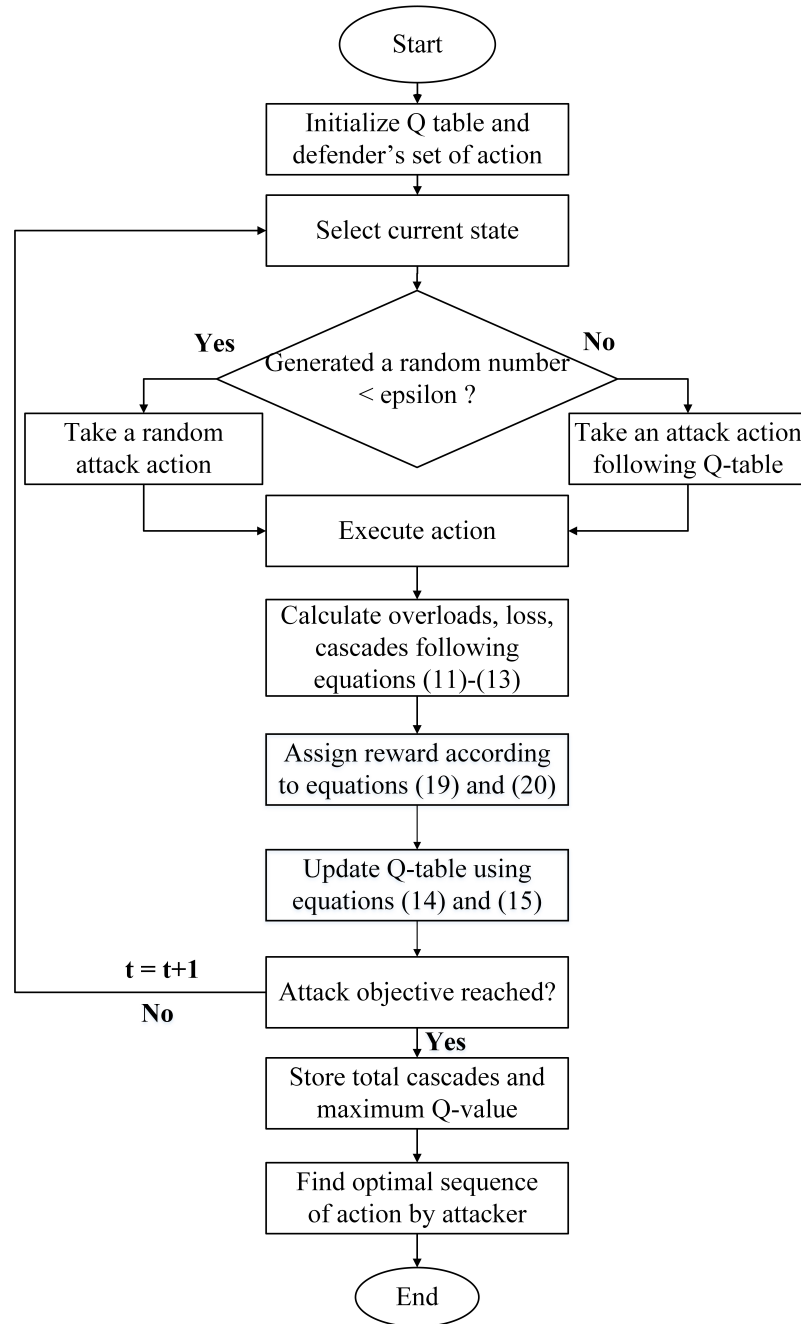


Figure 3.3. Proposed reinforcement learning algorithm to solve the attacker-defender two-person game in power system.

The final value of epsilon ( $\epsilon_f$ ) is a very small exploration probability for the agent to take random actions even at the end of the episodes. With the execution of the attack

action, a transmission line is switched to out of service condition if it is not protected by the defender. This action causes overloads in transmission lines (measured by equation (2.3)). Based on the re-dispatching of the generators from equations (2.1) and (2.2), load-shedding is determined. Then, generation loss and cascades are calculated, the reward is assigned, and the  $Q$ -table is updated. If the attack objective is reached at this time step, the game is over and  $Q$ -value is stored. If not, players go to the next time step ( $t = t + 1$ ), and the game is repeated until the attack objective is reached. With action,  $a_t$  taken at time step  $t$ , the action is executed in the power system environment, and the system state is updated including the failed transmission lines. During the game, a pair of actions  $(a, d)$  will lead to an immediate expected reward for both players. For the attacker, transmission line failures of the power system determine the rewards, denoted by  $R^A(s, a, d)$ . The attacker intends to maximize the total reward, while the defender tries to minimize the power system failures and minimize the total reward. Thus, the defender's expected reward is the negative of the attacker's reward. The defender's reward is denoted by  $R^D$  and  $R^D(s, a, d) = -R^A(s, a, d)$ . To this end, the smart grid attacker-defender game falls exactly into the two-player zero-sum game.

**Design of reward and Q-function** We consider only deterministic cases while implementing reinforcement learning. From the attacker's perspective, we define the  $Q$ -value as

$$Q_{att}(s, a, d) = R_A(s, a, d) + \gamma \sum_{s' \in S} V_{att}(s') \quad (3.30)$$

and the corresponding value function is defined as

$$V_{att}(s) = \max_{a \in A} \min_{d \in D} \sum_{a \in A} \sum_{d \in D} Q_{att}(s, a, d) \quad (3.31)$$

where  $V_{att}$  is denoted as the attacker's expected long-term reward for optimal strategies for the initial state  $s$  and  $Q_{att}(s, a, d)$  as the expected long term rewards for taking action  $a$  against defender's taken action  $d$ . Similarly, the defender will update  $Q$ -value as

$$Q_{def}(s, a, d) = R_D(s, a, d) + \gamma \sum_{s' \in S} V_{def}(s') \quad (3.32)$$

and corresponding value function as

$$V_{def}(s) = \min_{d \in D} \max_{a \in A} \sum_{a \in A} \sum_{d \in D} Q_{def}(s, a, d) \quad (3.33)$$

Here  $Q_{def}(s, a, d)$  is the expected future reward for the defender for adopting action  $d$  against attacker's action  $a$  for state  $s$ .

The learning player (attacker) is leading the learning procedure; while the opponent (defender) is a passive player, whose policy is pre-defined throughout the learning procedure. Considering that, the defender's strategy can be evolved from the attacker's progressive action strategy [139]. The main target of this work is to find the defender's critical action set. So, first the attacker is trained to find the optimal action strategy in the presence of a passive defender. Then, the critical transmission lines (frequently used transmission lines in the optimal strategies) are added to the defender's protection set.

The states of the game are presented as a combination of the status of all the transmission lines of the smart power grid, and it is denoted by  $S$ . For each state  $s$ , the status of the transmission lines are represented by binary number "1" or "0" as

$$s_t(l) = \begin{cases} 1, & \text{if line } l \text{ is in-service at time } t \\ 0, & \text{if line } l \text{ is out-of-service at time } t \end{cases} \quad (3.34)$$

If total number of transmission lines in the system is  $N$ , the state space can be described as  $S = \{s_1, s_2, s_3, \dots, s_{N_s}\}$  as shown in Figure 3.4. Initial state represents the original normal

condition of power system. Next steady state represents post attack condition of the power system. In this work, we consider deterministic transition.

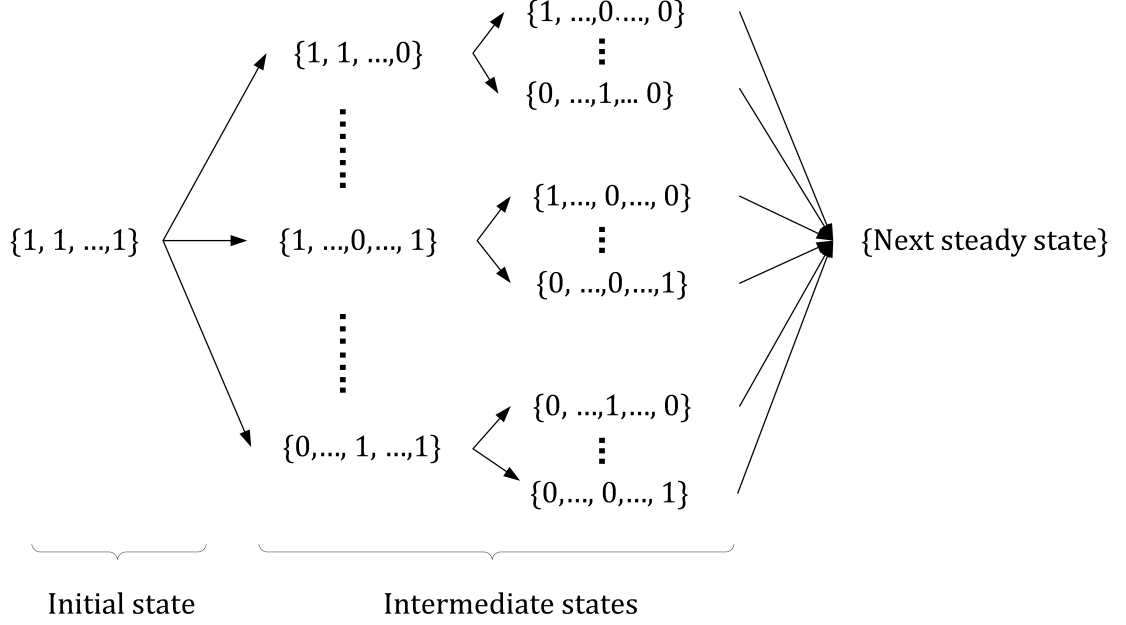


Figure 3.4. Transition from initial state to the new steady state through different intermediate states.

In this zero-sum game, the reward is assigned opposite to each other. So, if the number of instant transmission line outages for a game is  $N_0$ , then the reward is assigned following the conditions below

$$R^A(s, a, d) = \begin{cases} +1, & \text{if } N_0 \geq N_\theta \text{ and } k < N_\theta \\ 0, & \text{otherwise} \end{cases} \quad (3.35)$$

and

$$R^D(s, a, d) = \begin{cases} -1, & \text{if } N_0 \geq N_\theta \text{ and } k \geq N_\theta \\ 0, & \text{otherwise} \end{cases} \quad (3.36)$$

where  $N_\theta$  represents the attack objective and  $k$  is the number of actions taken by the attacker.

### 3.3 Numerical results and simulations studies

#### 3.3.1 Stage-game in adversarial environment

##### 3.3.1.1 Simulation study: linear programming

In this section, we analyze the consequences of the attack on IEEE 30 bus system. According to the assumptions made, there are 6 insecure targets ( $z_1, z_2 \dots z_6$ ) and the attacker is capable of attacking two of the targets at the same time. So, the final targets are the 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.  $z_1 z_2, z_1 z_3 \dots etc.$ ). 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 3.4 are the results of different attack and defend strategies (considering both player's action). In this section we analyze the results for IEEE 30 bus system. Table 3.4 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 3.1. 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 3.15 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  $w = [0.0305 \ 0.1988 \ 0.0585 \ 0 \ 0 \ 0 \ 0.6530 \ 0 \ 0 \ 0 \ 0 \ 0.0593 \ 0 \ 0]$ .

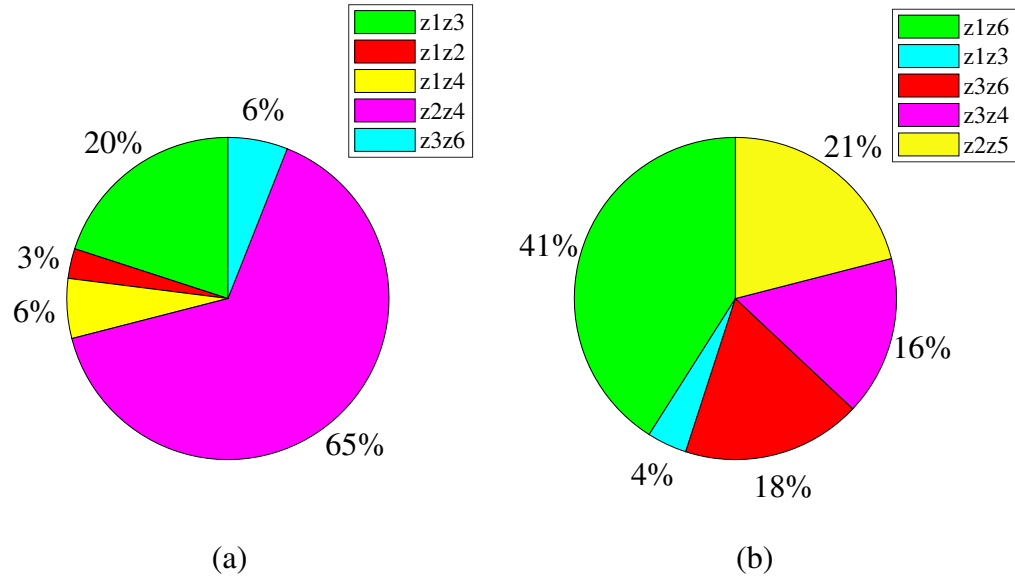


Figure 3.5. (a) Attacker's and (b) defender's mixed strategies for different targets in attacker-defender zero-sum two-player game for IEEE 30 bus system

Figure 3.5 (a) and (b) show the proportion of times that the attacker and the defender should attack and defend different targets. In figure 3.5(a), 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 3.5(b) 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.

### 3.3.1.2 Simulation study: reinforcement learning

Reinforcement learning can solve the one-shot game following the formulas from section 3.2.1.2. The simulation results are given and explained here in different case studies. First we apply deterministic Q-learning algorithm to conduct the game. Then we formulate the game for alternative action choices.

**Case study 1 (deterministic game):** W & W 6 bus system is considered as the test system for the game in this case study.

Table 3.5. Parameter information for the two-player zero-sum game between attacker and defender in W & W 6 bus system.

Parameter	Values
Test Case	W & W 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, $\gamma$	0.9
Epsilon, $\epsilon$	0.4
Total iterations	1000

Table 3.5 gives the value of the parameters considered for game formulation and simulation in 6 bus system. Here, epsilon,  $\epsilon$  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.

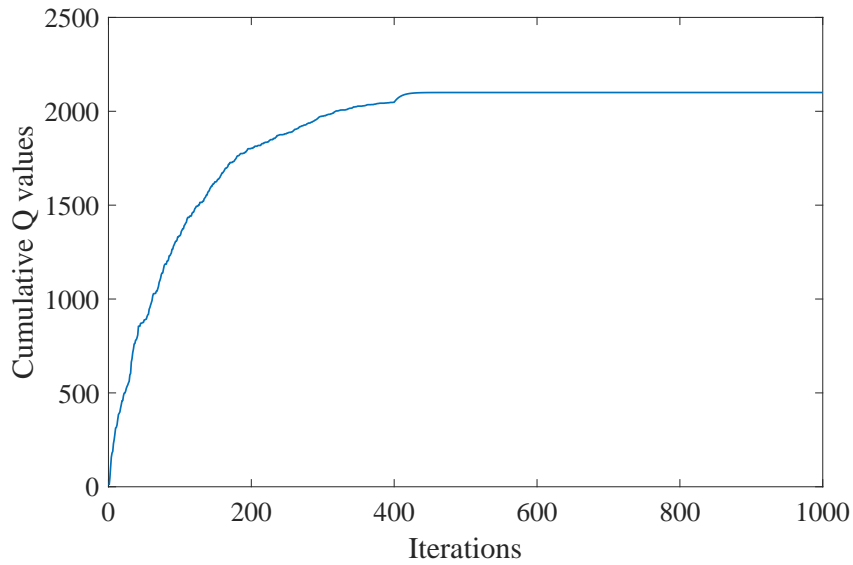


Figure 3.6. Attacker's cumulative Q-value per iteration in the adversarial two-player zero-sum game for W & W 6 bus system (average of 10 runs).

Figure 3.6 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 W & W 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.

**Case study 2 (game with alternative action choices for W & W 6 bus system):**

Now we conduct the game to have probabilities for alternative action selection for the adversaries. In this case study we assume, the defender is defending transmission line 1. We assume that the defended transmission line cannot be attacked.

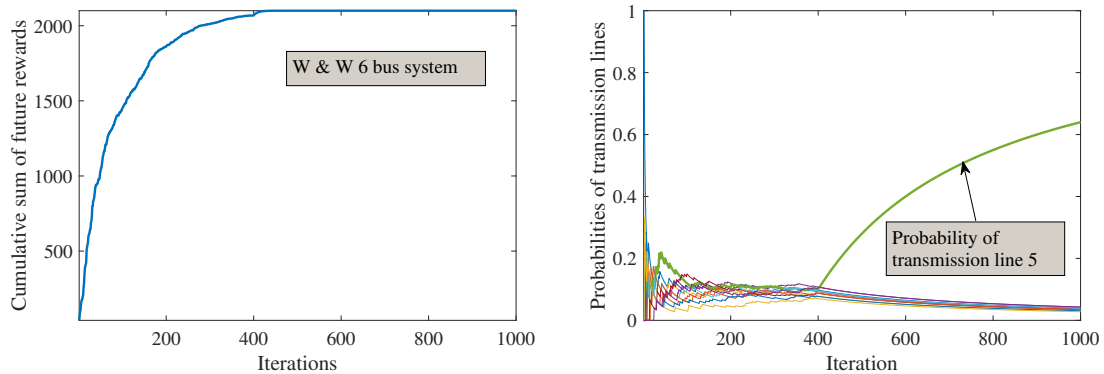


Figure 3.7. (a) Cumulative sum of future rewards of the attacker in the adversarial stage game (defender's protection set  $\{1\}$ ) (b) probability update for all the transmission lines to be selected as an action.

Figure 3.7(a) shows the cumulative sum of future rewards (Q-values) for the adversarial stage game conducted between the adversaries. The attacker conducts the game for 1000 iterations for learning through trial and error process. From the figure, we can see that after 400 iterations the learning agent converges to its optimal policy. To update the probabilities

of these action selection, the agent follows equation (4.30).

Figure 3.7(b) shows the probability update of the transmission lines of W & W 6 bus system to be selected as an attack action. The green curve is showing the probability update of transmission line 5. The initial oscillation in probability updating of the transmission lines represents the random exploration of the action selection by the attacker. From this figure we can see that, the probability of transmission line 5 to be selected as an attack action increases after enough exploration (after 400 iterations). The initial oscillations of the probabilities of the transmission lines happens due to the random action selection of the agents during the learning process (exploration). On the other hand, the probabilities of the other transmission lines to be selected as attack actions (optimal actions) drop while following the greedy policy.

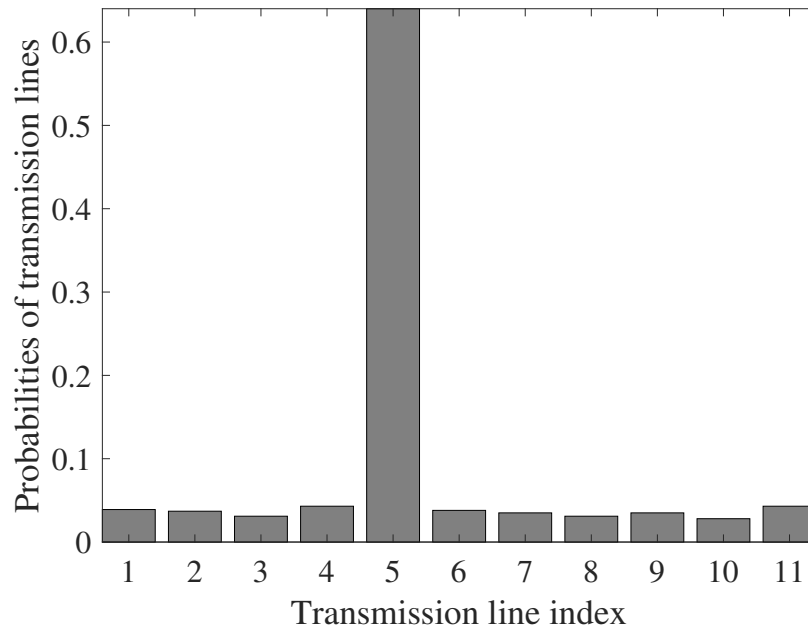


Figure 3.8. Action selection probabilities for the transmission lines of W & W 6 bus system.

Figure 3.8 shows the probabilities of the transmission lines for action selection. From

this figure, we can see that transmission line 5 has the highest probability to be selected as an attack action, while the defender is defending transmission line 1. In practical life, it might happen that after triggering the attack the transmission line is reinforced due to the protection scheme of the system or that specific transmission line or area is connected to distributed energy resources (DER). In that case, attacking on that transmission line will not be successful. In any of these cases, if transmission line 5 is not accessible, connected to DER or reinforced, the attacker will attack the transmission line with next highest probability. If transmission line 5 is inaccessible, the next highest probability goes with transmission line 4 or 11 which is 0.043. So, the attacker will select these transmission lines as the attack action. Next, we consider that learning the most critical transmission line of W & W 6 bus system, we adjust the defense policy of the defender. Now, the defender will protect transmission line 5. With this new defense policy, we conduct the game again.

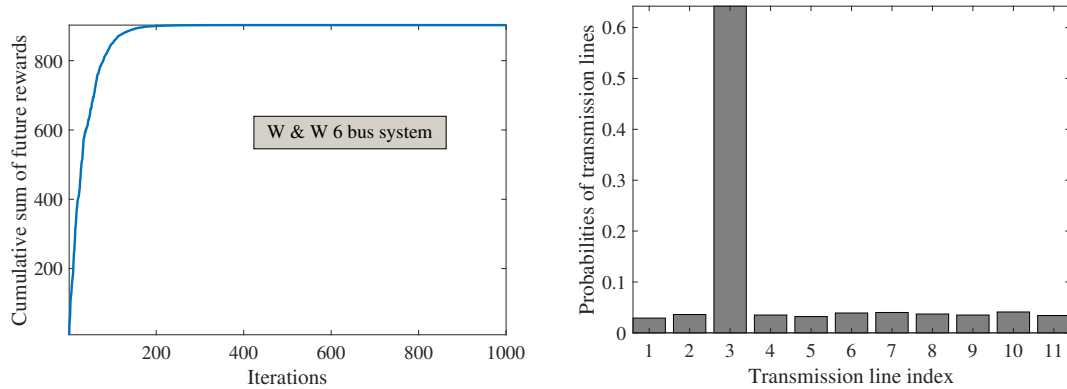


Figure 3.9. (a) Cumulative sum of future rewards and (b) action selection probabilities for the transmission lines of W & W 6 bus system with the adjusted defender's policy.

Figure 3.9(a) shows the cumulative sum of future rewards for the attacker. Compared to Figure 3.7(a), the adjusted policy has lower value of cumulative sum of future rewards (generation loss is reduced).

Figure 3.9(b) shows the action selection probabilities of the transmission lines of W & W 6 bus system with the adjusted defender's policy. So, according to this figure, while the defender is defending transmission line 5, the attacker will attack transmission line 3.

**Case study 3 (IEEE 39 bus system - alternative action choices and attack impacts):**

In this subsection, we use IEEE 39 bus as the test system to conduct the adversarial stage game, and then we analyze the impact of the attack on a simulated power system using PowerWorld simulator. First, we conduct the stage game on IEEE 39 bus system to identify the most vulnerable branch/branches from the system.

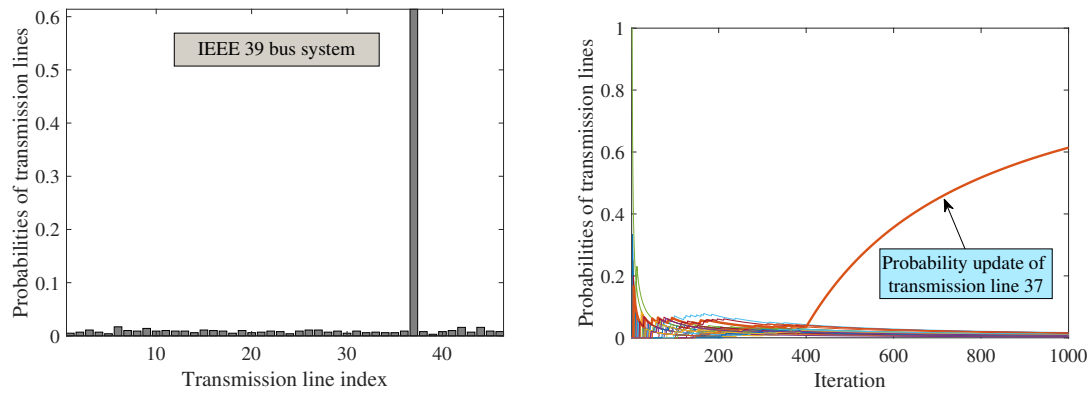


Figure 3.10. (a) Action selection probabilities and (b) probability update of the attacker's action selection probabilities for IEEE 39 bus system. The randomly predefined defense action is transmission line 5.

Figure 3.10(a) shows the action selection probabilities for adversarial stage game in IEEE 39 bus system. From this figure, we can see that transmission line 37 has the highest probability to be selected as an action by the attacker. From several runs, we found that there are actually three transmission lines that are most vulnerable in IEEE 39 bus system. They are transmission line 8, 12, and 37. And the selection probability of this transmission lines is 0.614.

Figure 3.10(b) shows the updating of the probabilities via trial and error in the learning process. The red curve shows the probability update of transmission line 37 to be selected as the attack action. Switching transmission line 37 will cause cascaded failure of several transmission lines (lines 12, 8, 5, 21, 18, 4, and 31). This attack also creates multiple sub-grids from different buses, such as bus 10, 14, 25, and 31 divides into 2, 3, 4, and 5 sub-grids, respectively. There are some other buses that divides into multiple sub-grids. Now we move forward to the PowerWorld simulator to observe the impact of the transmission line attack in the simulated power system. We conduct the simulation for 3 seconds and we assume the attack happens at 1.5 second. We attack transmission line 37 as the target. Transmission line 37 is connecting bus 6 and bus 31. Due to the attack, transmission line 6 opens from the both end. For evaluation of the impact we are considering voltage violation as the index. Several references reported different voltage limit [140], [141]. We assume, the voltage limit for bus voltages in per unit is 0.9 to 1.1.

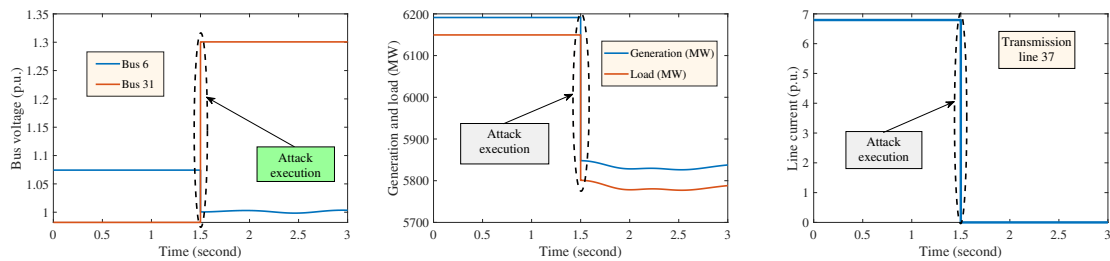


Figure 3.11. (a) Bus voltages at bus 6 and bus 31, (b) total generation and load change before and after the attack and (c) line current at transmission line 37 of IEEE 39 bus system.

Figure 3.11(a) shows the bus voltages at bus 6 and bus 31. Voltage at bus 6 drops from 1.074 to 1.003 which is within the range. But, voltage at bus 31 rises from 0.982 to 1.301 which is above the upper limit of the voltages (p.u.). So, bus 31 violates the voltage

limit. Figure 3.11(b) shows the generation and load change before and after switching transmission line 37 in IEEE 39 bus system. The attack over transmission line 37 reduces the generation and load of the system by creating disturbances in the system. Figure 3.11(c) shows the line current flow (p.u.) in transmission line 37. The figure shows that, the current drops to zero at 1.5 second due to the initiation of the attack.

Apart from the impacts on the bus voltages and generations, frequency of the system also become unstable due to the attacks. According to the European Network of Transmission System Operators for Electricity (ENTSOE) standards, if the frequency of the European grid goes beyond the range  $[47.5Hz - 51.5Hz]$ , the system can hardly avoid a blackout.

### 3.3.2 Multistage sequential game

The simulation is conducted using Matlab *R2018a* on a standard PC with an Intel(R) Core(TM) *i7 – 6700* CPU running at 3.40GHz and 24.0 GB RAM. There are three loops in our simulation: *runs*, *episodes*, and *iterations*. The *iterations* loop ends when the given number of actions are used by the attacker. The *episodes* loop is the main loop in which the attacker and the defender interact to learn the optimal policy. As the number of episodes is increasing, the attacker tends to learn the optimal policy. At the end of the episodes, the attacker and the defender reach the Nash equilibrium point. A number of runs are conducted to find out different Nash equilibrium points of the game. So, the whole game simulation is conducted for several runs. Each run consists of a number of episodes. And every episode consists of a number of iterations. The number of iterations represents the number of actions taken by the attacker. The number of episodes is the required number of trials for the agent in the learning process. The Q-table is defined with the attacker's

state-action pairs and Q values. The discount factor,  $\gamma$ , is applicable only for the attacker's learning process. The attacker decides the value of  $\gamma$ . In the simulation studies, both the attacker and the defender have a limited number of actions. The attacker can only attack one transmission line at a time. The number of actions taken is limited to the targeted number of transmission line outages. In other words, the attacker's actions cannot outnumber the attack objective of line outage. In the following subsections, we will present the two-player zero-sum game between the attacker and the defender on the W & W 6 bus system and the IEEE 39 bus system.

**Case I : Study of branch outage for W & W 6 Bus System** In this case, we investigate the two-player zero-sum game behavior between the attacker and the defender for W & W 6 bus system. This case study gives us a step-by-step understanding of the dynamic game decision-making process and the achieved optimal attack sequences.

Table 3.6. Parameter information for W & W 6 bus system.

Parameter	Value
Total number of branches, N	11
Total generation	210 MW
Attack objective, $N_\theta$	30% line outage (minimum 4)
Defender's action set (line number)	$\langle 1, 2 \rangle$
Discount factor, $\gamma$	0.9
Maximum iteration per episode	100
Total number of episodes	600
Total number of run	100
Exploration probability	0.8
Final epsilon, $\epsilon_f$	0.003
Epsilon divided	Every 10 episode
Epsilon divided by	1.1

Table 3.6 gives the parameter settings for the attacker-defender game for the W & W 6 bus system. Initially, the defender's protection set  $\langle 1, 2 \rangle$  is used, meaning that the de-

fender will protect these lines, which will not be switched when the attacker tries to attack. However, the protected transmission lines will still be switched to out-of-service during the cascading failure process. There are 100 independent runs for this simulation and each run consists 600 episodes. The exploration probability starts with 0.8 and gradually drops to a small value  $\varepsilon_f$  until the end. For example, the  $\varepsilon$  will be divided by 1.1 every 10 episodes in this case. A small  $\varepsilon_f$  will ensure the convergence to the optimal policy and still avoid the local minimum point. The computational time for one run of this game is 8.64 seconds.

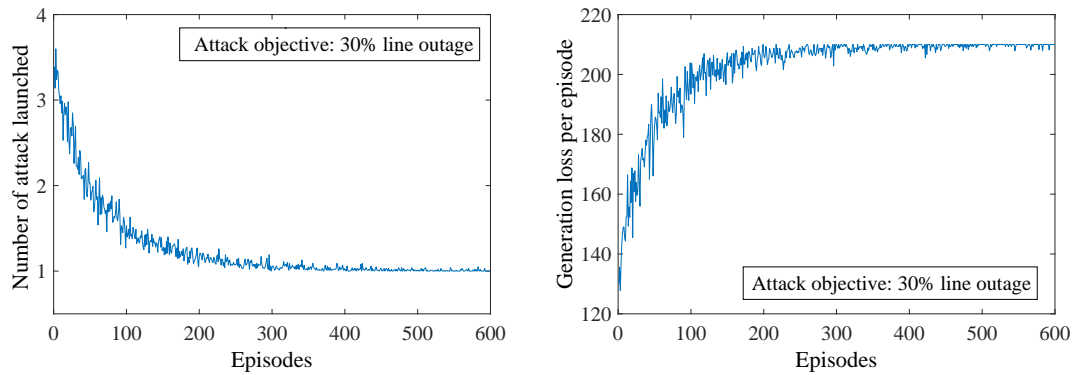


Figure 3.12. (a) Convergence of the attacker's number of actions and (b) generation loss due to the attacker's action in W & W 6 bus system (average of 100 runs). The defender's protection set is  $\langle 1, 2 \rangle$ .

**Convergence of attack sequences and generation loss:** Figure 3.12(a) shows the convergence curve for the attacker's number of actions to reach the Nash equilibrium point. The number of attack actions in the Y axis are non-integer sometimes because of the average of 100 runs. Among 100 independent runs, all of them converged to the optimal policy (minimum number of actions). We found that the attacker needs to take only one action (attacking line 5) to reach the line outage threshold (30%). Figure 3.12(b) shows the convergence curve for generation loss. Attacking line 5 causes the maximum generation loss

of 210 MW, which is also the total generation capacity of the system. From the results, the power system operator is suggested to first protect line 5, which is the most critical line in this benchmark system. From Figures 3.12(a) and 3.12(b), we can notice oscillations at the end of the episodes. The value of  $\epsilon_f$  creates a few random actions at the end of episodes.

**Case II: Study of impact of exploration variation** In this case study, we show the impact of exploration variation on the percentage of convergence to optimality. We also show how timely adjustment of the defender's action policy from the learned attacker's action policy improves the security of the defense scheme. IEEE 9 bus system and IEEE 39 bus system are considered as the test systems for this case study. First, we start with 20% exploration and then we increase the exploration rate up to 80%. The required number of episodes for IEEE 9 bus system and IEEE 39 bus system are 5,000 and 8,000, respectively. We consider the attack objective is to cause a 30% of transmission line outage.

Table 3.7. 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%

Table 3.8. 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%

Table 3.7 and 3.8 shows the number total number of exploration required with the associated exploration rates.

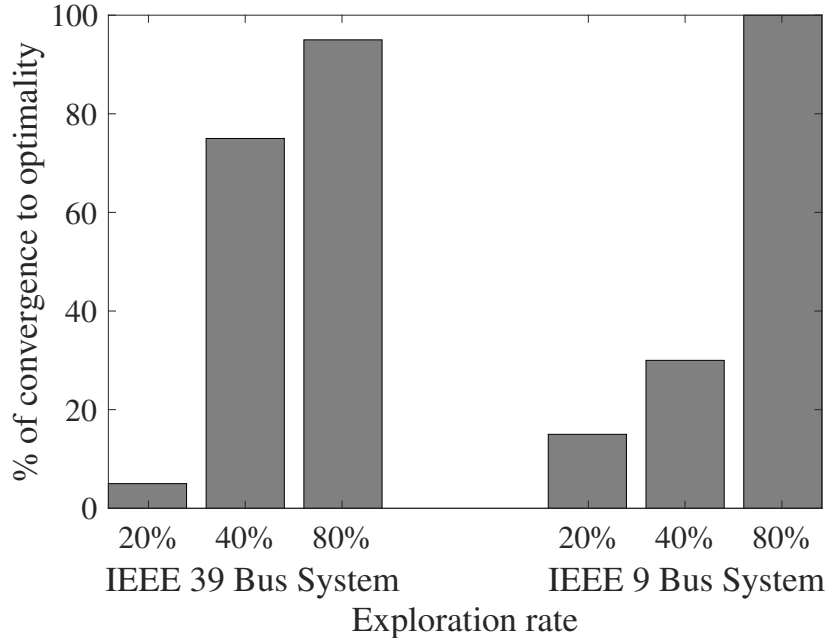


Figure 3.13. Percentage of convergence to optimality for IEEE 9 bus system and IEEE 39 bus system with different exploration rates.

Figure 3.13 shows the percentage of convergence to optimality of the attacker's action policies with different exploration rates. We can see that the convergence percentage increases with increased exploration rate. For example, in the case of IEEE 39 bus system with 20% exploration probability, only 5% attack strategies reach optimality. But as we increase the exploration probability to 40% and 80%, the percentage of convergence to optimality increases up to 75% and 95%, respectively. Now let us consider a 30% transmission line outage as the attack objective, the defender's random predefined action set as  $\{1, 2, 3\}$ . With this action set, the attacker's learned policy detects that the transmission lines 8, 27, and 37 are the most frequently selected targets. So, we gradually start including these transmission lines to the defender's policy to show how the trained policies of the

attacker help the defender make the system security stronger.

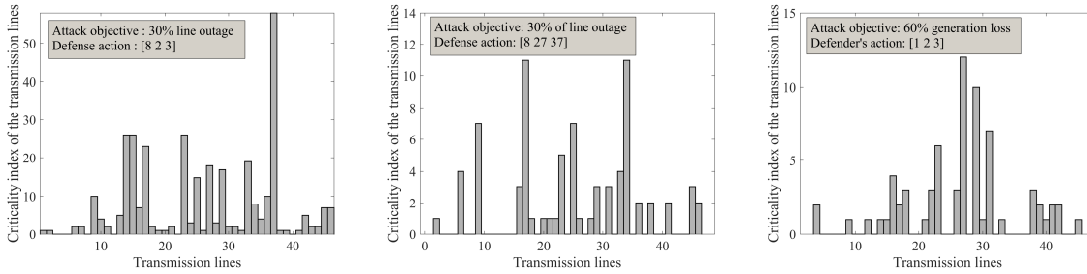


Figure 3.14. Criticality index of the transmission lines of IEEE 39 bus system to be selected as the attack actions by the attacker when the defender defends transmission line (a)  $\{8, 2, 3\}$ . The number of unique optimal policies here is 82. (b)  $\{8, 27, 37\}$ . The number of unique optimal policies here is 19. (c) Criticality index of the transmission lines of IEEE 39 bus system to be selected as the attack action by the attacker. The attack objective is to cause 60% generation loss. The defender defends transmission line  $\{1, 2, 3\}$ . The number of unique optimal policies here is 35.

Figure 3.14(a) shows the criticality index of the transmission lines of IEEE 39 bus system to be selected as attack actions. The criticality index represents the frequency of the transmission lines to be selected as the attack actions. The defender is defending transmission lines 8, 2, and 3. The number of unique optimal policies reduces from 97 to 82 out of 100 runs. Now we include all the three transmission lines (8, 27, and 37) to the defender's protection set and observe how the number of unique optimal policies changes. Figure 3.14(b) shows the frequencies of the transmission lines from IEEE 39 bus system for unique optimal policies of the attacker. The defender is defending transmission lines 8, 27, and 37 and the number of unique optimal policies of the attacker reduces to 19 out of 100 independent runs.

Now we change the attack objective from transmission line outage to generation loss. So, with 60% generation loss as the attack objective, we randomly select the defender's defense action as  $\{1, 2, 3\}$ , and we found that the transmission lines 23, 27, and 29 are the

most frequently used transmission lines as attack actions out of 100 independent runs. If we take the unique optimal policies (35 out of 100), we found transmission lines 27, 29, and 31 are the mostly used transmission lines as attack actions, hence the most critical ones.

Figure 3.14(c) shows the criticality index of the transmission lines of IEEE 39 bus system when the defender's action is randomly pre-defined (1, 2, and 3). Similar to the previous attack objective, protecting the frequently used transmission lines (by including in the defender's protection set), the number of optimal action sequences can be reduced.

This case study shows that, with increased exploration probability, the learning agent can achieve higher percentage of convergence to optimality. On the other hand, this case study also proves that, with the identified critical transmission lines from the attacker's policy it is possible to adjust the defender's policy and make the defense strategy stronger.

**Case III : Study of branch outage for IEEE 39 bus system** In this case, we test our proposed algorithm on IEEE 39 bus system. First, we identify the critical transmission lines through a pre-defined protection set. Then we include the critical transmission lines into the protection set, test the attacker's learning performance, and validate the protection effect. IEEE 39 bus system has 46 branches and its total generation capacity is 6150 MW. We pre-define the defender's action set as  $D = \langle 1, 2, 3 \rangle$ .

Table 3.9. Parameter information for IEEE 39 bus system.

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

Table 3.9 gives the parameter values for IEEE 39 bus system to conduct the simulations. The attacker starts with random actions and faces the defender's protection set at the beginning. There are 100 independent runs in total, and the number of episodes is 8000 per run. The value of exploration probability  $\varepsilon$  starts from 0.8 and decreases every 100 episodes by dividing by 1.1. In this case, the attack objective 30% line outage means the 14 transmission line outages.

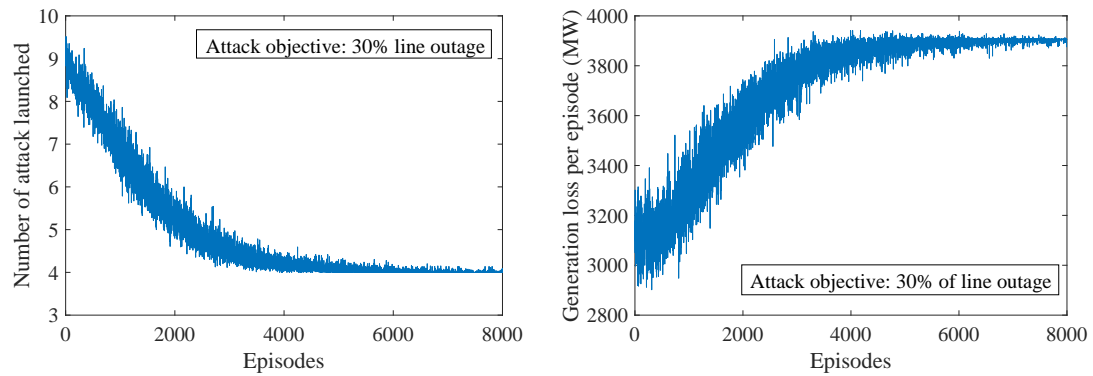


Figure 3.15. Convergence of (a) the attacker's number of actions and (b) generation loss due to the attacker's action for IEEE 39 bus system (average of 100 runs). The defender's protection set is  $\langle 1, 2, 3 \rangle$ .

**Convergence of attack sequences and generation loss:** Figure 3.15(a) shows the convergence to the optimal number of attack actions. We can see, after enough learning and exploration, the attacker reaches the optimal number of attack actions.

Table 3.10. The attacker's action sequences for IEEE 39 bus system.

Run index	Number of actions	Action sequence	Total outages (including cascaded outages)	Generation loss (MW)
1	4	$\langle 8, 29, 10, 13 \rangle$	$\langle 1, 3, 7, 8, 10, 12, 13, 18, 19, 21, 22, 23, 29, 31 \rangle$	3649.56
2	4	$\langle 36, 27, 8, 32 \rangle$	$\langle 1, 3, 7, 8, 12, 17, 18, 19, 22, 27, 29, 31, 32, 36 \rangle$	3533.44
3	4	$\langle 37, 23, 14, 40 \rangle$	$\langle 3, 4, 5, 8, 12, 14, 18, 21, 22, 23, 29, 31, 37, 40 \rangle$	3584.53
4	4	$\langle 36, 27, 8, 28 \rangle$	$\langle 3, 4, 5, 8, 12, 15, 18, 21, 22, 23, 29, 31, 34, 37 \rangle$	3177.46
...	...	...	...	...
...	...	...	...	...
99	4	$\langle 42, 8, 27, 45 \rangle$	$\langle 3, 7, 8, 12, 18, 19, 21, 22, 23, 29, 31, 33, 34, 38 \rangle$	4212.18
100	4	$\langle 37, 23, 32, 15 \rangle$	$\langle 3, 4, 5, 8, 12, 14, 18, 21, 22, 27, 29, 31, 34, 45 \rangle$	5582.04

In Table 3.10, there are many different policies to reach the objective (30% line outage) and the minimum number of attack actions in those policies is 4. Due to the space limitation, we could only show several attack action sequences among 100 runs. For example, the first action sequence in the table  $\langle 8, 29, 10, 13 \rangle$  gives the total outages of  $\langle 1, 3, 7, 8, 10, 12, 13, 18, 19, 21, 22, 23, 29, 31 \rangle$ .

**Comparison of multistage attack and one-shot attack:** To illustrate the importance of this dynamic game, let us compare the total cascaded outages between a multistage attack and a one-shot attack. This action sequence is shown in Table 3.11. In the upper portion of the table, the left column represents the attack actions taken and the right column represents the cascaded outages due to the actions. The same attack sequence is applied as the one-shot attack which is shown in the lower portion of Table 3.11. Using the same attack sequence, multi-stage attack gives 14 transmission line outages whereas one-shot attack gives 11 transmission line outages. So a multistage attack generates more line outages than a one-shot attack.

Table 3.11. Comparison of cascading failures between a multistage attack and a one-shot attack. This table shows the attack actions for optimal action sequence and cascaded outages for run 1 from Table 3.10.

Multistage attack	
Attack actions	Cascaded outages
8	12, 21, 18, 19, 3, 31, 7
29	23
10	22
13	1
One-shot attack	
Attack actions	Cascaded outages
8, 29, 10, 13	23, 7, 18, 19, 22, 3, 31

**Criticality analysis:** From Table 3.10, we plot the frequency of all the line attacks of 100 runs in Figure 3.16. Here, criticality index (y label) counts the number of the transmission lines attacked in 100 runs. From this game with the defender's action set as  $\langle 1, 2, 3 \rangle$ , we found 97 unique attacker's action sequences.

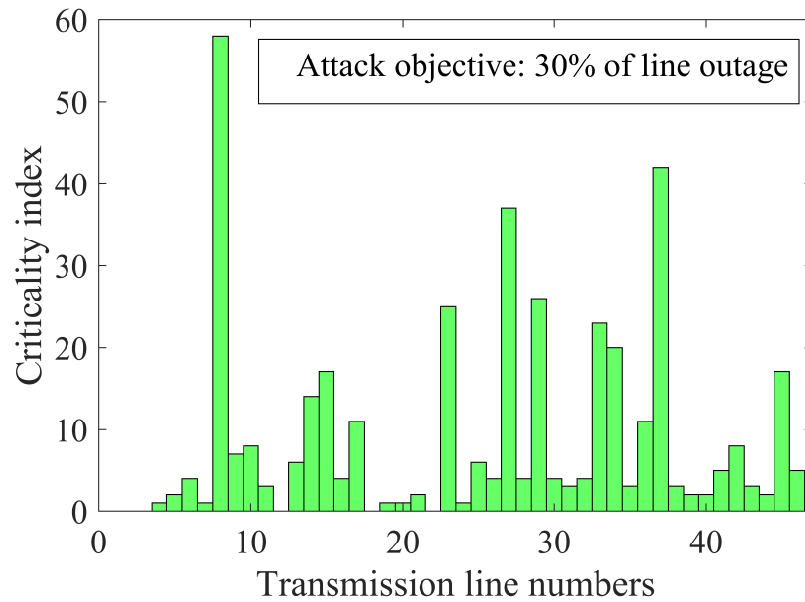


Figure 3.16. Frequency of the line attacks for IEEE 39 bus system over 100 independent runs. The defender's protection set is  $\langle 1, 2, 3 \rangle$ .

From Figure 3.16, we can see that, there are several branches (i.e., line numbers 8, 37, 27) appeared more times than the others. From the experience in Case I, these lines are the critical lines for the IEEE 39 bus system. In order to analyze the effect of the defender's policy in this game, we will apply these three critical transmission lines into the defender's protection set one by one.

Table 3.12. Number of converged and unique optimal strategies for the defender's different action's set for IEEE 39 bus system.

Defender's action	Attacker's no. of unique optimal strategies
$\langle 1, 2, 3 \rangle$	97
$\langle 8, 2, 3 \rangle$	82
$\langle 8, 37, 3 \rangle$	23
$\langle 8, 37, 27 \rangle$	19

Table 3.12 shows the number of the attacker's unique optimal strategies against the defender's protection strategies. First, the defender includes line 8 in its defense strategy; the defender's new defense strategy becomes  $\langle 8, 2, 3 \rangle$ . We conduct 100 independent runs and obtain that the unique converged attack policies reduces from 97 to 82. Second, we add transmission line 37 in the defender's action set and the new defense set is  $\langle 8, 37, 3 \rangle$ . The same game is conducted, 27 out of 100 runs converge to the optimal number of actions. There are 23 unique optimal action strategies among the 27 optimal converged strategies. Third, we add those three critical lines to the defender's defense set, which is now  $\langle 8, 27, 37 \rangle$ . We conduct the same game, and obtain 21 optimal actions sequences. There are 19 unique attack sequences. From the above settings and results, we observe that the critical lines are the ones with high chances of attack by "smart" attackers, and

will also cause severe cascading failures. The more critical lines we include in the defender's protection set, the fewer optimal policies will be available for the attacker to reach the objective with the minimum number of actions. The results provide useful suggestions for power system operators and government authorities to manage system planning and cyber/physical protections.

#### Case IV : Study of generation loss for IEEE 39 bus system

**Convergence of attack sequences:** Generation loss is also one of the important factors in the smart grid security. In this case, we use generation loss as the attack objective. Like case *III* the IEEE 39 bus system is used here as the benchmark system. The parameter settings are similar as those in Case III except the following: the attack objective here is to cause 3690 MW of generation loss (60% of 6150 MW), value of  $\varepsilon = 0.4$ ,  $\varepsilon$  is divided each 50 episode by 1.1, the total number of episode is 5000. We define the defender's defense set as  $\langle 1, 2, 3 \rangle$ . The attacker plays against this defender's defense set and reaches to its expected optimal policy.

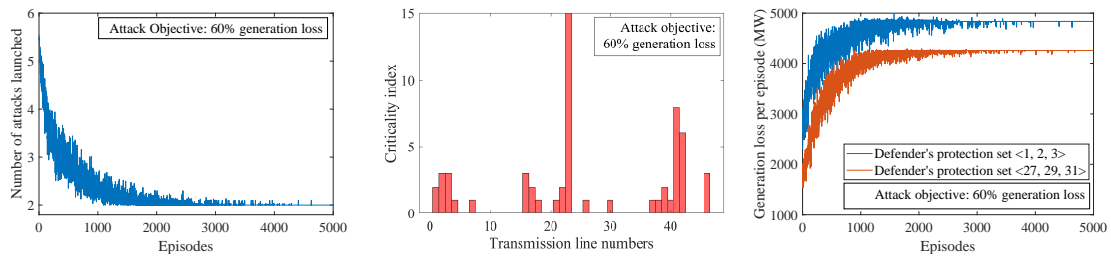


Figure 3.17. (a) Convergence of the attacker's number of actions (average of 100 runs). The defender's protection set is  $\langle 1, 2, 3 \rangle$ . (b) Frequency of the transmission line attacks for IEEE 39 bus system. The defender's protection set is  $\langle 27, 29, 31 \rangle$ . (c) Convergence of generation loss due to the attacker's action (average of 100 runs).

Figure 3.17(a) shows the convergence of the attacker's strategy to optimal policy for the

IEEE 39 bus system. From Figure 3.17(a), we can see that only two actions are required to achieve the attack objective. Among the 100 runs, there are 35 unique sequences. From the 35 learned unique strategies, we identify lines 27, 29 and 31 are the most critical lines in the system to reach 60% (3690 MW out of 6150 MW) generation loss. The computational time for one run of this game is 161.45 seconds.

**Criticality analysis:** For further investigation, we apply these three critical lines into the defender's protection strategy and observe the game behavior. With the defense set  $\langle 27, 29, 31 \rangle$ , 29 unique strategies out of 100 runs converge to the optimal number of actions which is shown in Figure 3.17(b). It indicates that including critical lines in the protection set will give the attacker fewer choices to reach the attack objectives. Also, after including the critical transmission lines in the defender's protection set, the average generation loss per episode decreases. Figure 3.17(c) shows the average generation loss for the defender's protection set  $\langle 1, 2, 3 \rangle$  and  $\langle 27, 29, 31 \rangle$ .

**Case study V: exploration of the post attack impact on power system** In this case study, we select any learned attack policy of the attacker. Then we apply this learned attack policy (optimal attack actions) on a simulated power system using PowerWorld simulator to analyze the impact of the learned attack policy. We insert the attack targets in the PowerWorld test case for disconnecting in timely manner. We record the result in the ram for future use. For simplicity, we conduct the attack on IEEE 39 bus system consecutively after 0.5 seconds. So, we conduct the simulation for 2.5 seconds. And trigger the attacks every 0.5 seconds.

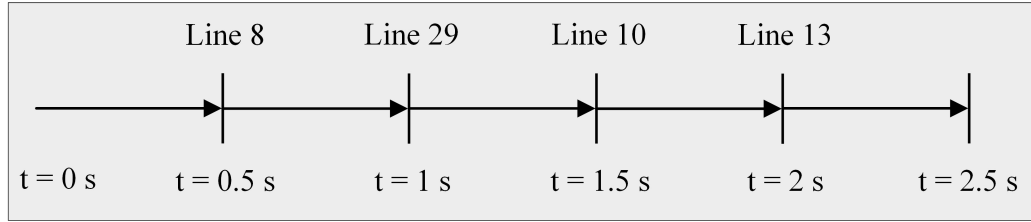


Figure 3.18. Time scale for the attack simulation on IEEE 39 bus system in PowerWorld simulator

Figure 3.18 shows the attack timescale for attack on IEEE 39 bus system in PowerWorld software. To assess the criticality of the learned attack policies, we consider voltage violation as the evaluation index. There are different references for voltage limits of the buses in per unit [142], [143]. We consider the range of the bus voltage is [0.9 to 1.1]. Exceeding 1.1, and dropping below 0.9 is considered as voltage violation. There will be some other lines, whose voltage will shoot to very close to the upper limit or the lower limit. Further disturbances in the system will cause them violate the voltage limitation.

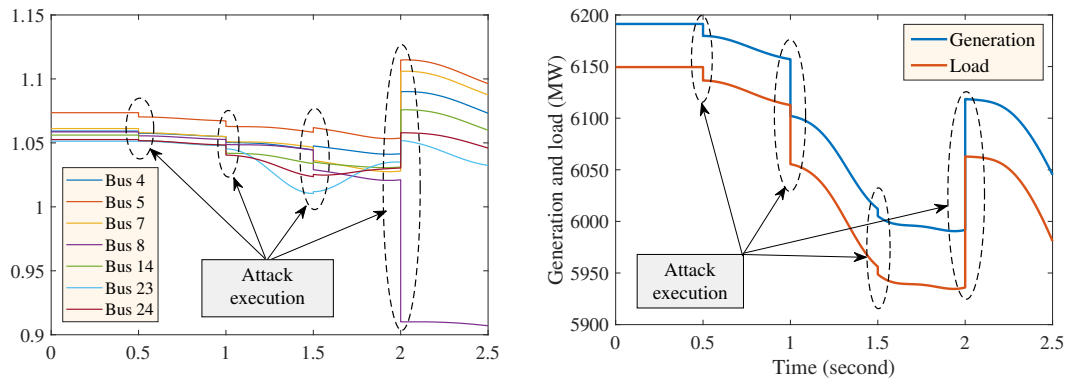


Figure 3.19. (a) Voltages (p.u.) at the buses connecting the target transmission lines. (b) Change in generation and load (in MW) due to the attacks in IEEE 39 bus system.

Figure 3.19(a) shows the voltages of the buses connected to the attacked transmission lines. From this figure we can see bus 5 and 7 exceeds the voltage (p.u.) upper limit (1.1 p.u.). Figure 3.19(b) shows the the changes in generation and load of the total system (IEEE

39 bus system). We can see the generation and load drop with each attack except the last attack. Due to the 4<sup>th</sup> attack the generation and load increases for a fractional seconds and then again drops.

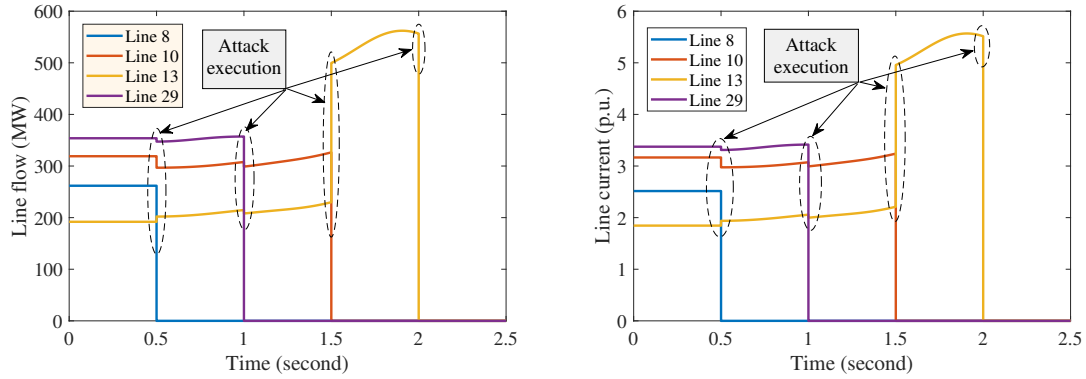


Figure 3.20. (a) Line flow and (b) line current of the attacked transmission lines (8, 10, 13, and 29).

Figure 3.20(a) shows the line flow through the attacked transmission lines. We can see that, with each transmission line switching attack, the value of the line flow drops to zero. Figure 3.20(b) shows the line current flowing through the attacked transmission lines. The value of the current drops to zero for each line associated with the optimal attack sequence. Both of the Figures 3.20(a) and 3.20(b) prove that the line switching attack is successful in switching the target transmission lines from active status to inactive status.

Executing line switching attack on the smart electric power transmission and distribution system creates a major impact on line currents, bus voltages, and generation. Voltage instability or voltage collapse is caused due to the inadequacy of the generated reactive power. This can happen because of the changes in the system configurations caused by transmission line outages, changes in the active or reactive power demand, or loss of generators. Apart from these impacts, these attacks can cause significant disturbance in the

frequency of the power system which can eventually leads to large scale blackout.

### 3.4 Summary

This chapter proposes a new solution for the two-person zero-sum attacker-defender multistage dynamic game (deterministic) in smart-grid security based on reinforcement learning. The learned attacker's optimal action sequence gives important information regarding critical transmission lines of a power system. Theoretical analysis of convergence is provided to ensure the effectiveness of the algorithm. Case I finds the critical transmission lines of the W & W 6 bus system for the defender's pre-defined action set for individual games. Case II gives us the critical transmission lines of the IEEE 39 bus system and further indicates that adjusting the defender's action from the attacker's learned critical transmission lines leaves the attacker fewer choices to make a successful attack. A comparison is shown in Case II indicates total transmission line outages caused by the multistage attack are bigger than a one-shot attack. A comparison of the defender's different strategies in Case III shows that adjusting the defender's strategy reduces the number of successful/optimal attacks, and that adjusting the defender's strategy reduces the average generation loss due to the attack. From these case studies, we can suggest that the learning of "simulated" attackers can eventually help the defender to better prepare defense strategies.

In this smart grid application, we notice that the theoretical number of states according to our formulation is given as  $2^N$ , where  $N$  is the number of transmission lines. Due to the reality of the system and the enforced constraints, many theoretical states do not exist and the attacker only needs to visit a limited number of states. For example, in W & W 6 bus system with 30% attack objective, the number of visited unique states is 37. In IEEE 39

bus system with 10% attack objective, the number of visited unique states is 13,130. The performance of identifying of vulnerable branches in the game could be further improved by introducing new deep learning and deep reinforcement learning techniques in the future.

## CHAPTER 4 Reinforcement learning based solution for repeated game

### NOMENCLATURE

$S$	Status of the transmission lines
$k$	Order of the contingency or attack
$t$	Attack timescale
$S_A$	Attacker's target set. A predefined set of power system transmission lines.
$S_D$	Defender's target set. A predefined set of power system transmission lines.
$G$	Contraction mapping.
$X$	Maximum time step.
$H$	Value iteration operator.
$N(S_A)$	Size of the attacker's target set.
$N(S_D)$	Size of the defender's target set.
$TLC$	Total loading capacity of the system (MW).
$C_w$	Cost of not attacking.
$C_a$	Cost of attacking (consumed resource from the system).
$C_m$	Cost of defending (holds the same property as $C_a$ ).
$B_A$	Attacker's allocated budget (defined by a number of actions).
$B_D$	Defender's allocated budget (holds the same property as $B_D$ ).
$\sigma$	The attacker's probability of attack or strength.
$\delta$	The defender's strength (the probability of defending).
$U_A$	Attacker's mixed strategy payoff.
$U_D$	Defender's mixed strategy payoff (holds the same property as the $U_A$ ).
$U_a$	The attacker's immediate mixed strategy payoffs in the repetitions.
$U_d$	The defender's immediate mixed strategy payoffs in the repetitions.
$C_i$	Immediate damage caused by the attack, the unit is in MW.
$P_a$	Total loss by the attack (MW).
$\epsilon$	Exploration rate.

$\gamma$	Discount factor of long term or short term reward.
$F$	Total number of runs in the game.
$Q$	Quality of the state $s$ , associated with action $a$ and $d$ .
$T$	Transition function between the states.
$V_a(s')$	Value at the next state $s'$ for the attacker.
$V_d(s')$	Value at the next state $s'$ for the defender.
$Pr$	Probability of a state-action pair $(s, a, d)$ .
$C$	Total number of times the state $s$ is visited for the specific action $a$ , and $d$ .
$\pi$	Action selection probability.
$\Pi$	Repeated game function.
$R$	Reward assigned for action $a$ and $d$ .
$Rep$	Repetition of a repeated game.
$a$	Action taken by the attacker.
$d$	Action taken by the defender.
$Z$	Represents active status of a line $l$ at time $t$ .

#### 4.1 Chapter overview

Due to the rapidly expanding complexity of the cyber-physical power system, the probability of system malfunctioning or failing is increasing. Most of the existing works combining smart grid (SG) security and game theory fail to replicate the adversarial events in the simulated environment close to the real life events. In this chapter, a repeated game is formulated to mimic the real life interactions between the adversaries of the modern electric power system. First, we formulate the environment for the attacker-defender interaction in the smart power grid. We provide a strategic analysis of the attacker-defender strategic interaction using a game theoretic approach. We apply repeated game to formulate the problem, implement it in the power system, and investigate for optimal strategic behavior in terms of mixed strategies of the players. In order to define the utility or cost function for the game payoffs calculation, generation power is used. Attack-defense budget is also incorporated with the attacker-defender repeated game to reflect a more realistic scenario. A comparison between the proposed game model and the All monitoring model is provided to validate the observations.

Then, the optimal action strategies for different environment settings are analyzed. The advantage of the repeated game is that the players can generate actions independent of the previous action history. The solution of the game is proposed based on reinforcement learning, which ensures the desired outcome in favor of the players. The outcome in favor of a player means achieving higher mixed strategy payoff compared to the other player. Different from the existing game theoretic approaches, both the attacker and the defender participate actively in the game and learn sequences of actions applying to the power trans-

mission lines. In this game, we consider several factors (such as, attack and defense costs, allocated budgets, and the players' strengths, etc.) that could effect the outcome of the game. These considerations make the game close to the real life events. To evaluate the game outcome, both players' utilities are compared and they reflect how much power is lost due to the attacks and how much power is saved due to the defenses. The players' favorable outcome is achieved for different attack and defense strengths (probabilities) of the players. IEEE 39 bus system is used here as the test benchmark. Learned attack and defense strategies are applied on a simulated power system environment (PowerWorld) to illustrate the post-attack effects on the system.

Due to the modernization of CPPS, humans/bots with heinous intention can exploit the system to access the CPPS and create damage to the system. To explain this behavior of the human/bot attackers, game theory is used. Game theory is able to interpret the complex interaction of the rational attacker-defender interaction with the help of mathematics and rational assumptions. Game theory is applied to the power system security to understand this complex behavior. Various algorithms are proposed, implemented, and used to solve the complexity of the behavior and provide a solution in response to the players' action [114], [115].

Several game theoretic approaches have been used in the domain of power system security. In [110], authors proposed a game-theoretic framework for the analysis of cyber-switching attacks and control-based mitigation in cyber-physical power system. In [58], [126], [144], authors applied a static game to investigate the vulnerabilities in power system. In [124], authors introduced a two-player zero-sum game between the adversary and the defender, and evaluated the equilibrium of defense mechanisms under network config-

urations.

All these game theoretic approaches need to have access to information regarding the opponent players' previous action history. Also, in some cases, information might be missing due to the environmental settings of the game. In that case, the game theoretic behavior may not interpret the actual complexity and solution.

Machine learning methods are becoming popular along with the increasing complexity of the critical infrastructures and their exposure to the security threats. Several machine learning (ML) algorithms are being used for detecting events and anomalies, finding critical elements of a power system, and implementing the interactions between the adversaries in a power system. These machine learning algorithms include classification and clustering methods, reinforcement learning algorithms, and artificial neural network, etc [16]–[18]. Securing the modern electric power grid is one of the most challenging issues nowadays. The modern electric power grid or the smart grid (SG) is an interconnected network of large number of heterogeneous devices. This complex interconnection of SG exposes the whole network to severe security threats. High integration of information and communication technology (cyber-layer) with the SG (physical layer) results in a complex and efficient cyber-physical power system (CPPS) [145]–[147]. The motivation of the attackers to target the CPPS includes causing economic downfall, affecting large group of individuals, damaging valuable infrastructures, and causing political damages. Incidents in Nigeria, Ukraine, the United States, and Kenya are some of the most recent power outage events around the globe [148]–[150]. According to the annual report of the United States by EATON, the total number of events related to the power outage that happened in the United States was 3,526 and they affected almost 36.7 million people in 2017. The total

duration of these outages was 2,84,086 minutes which are approximately equivalent to 197 days [12]. Cyber-attack is one of the major concern which is responsible for causing huge damage in the electric power grid. Dragonfly, Nightmare, Industroyer, Stuxnet, Wannacry, NotPetya, etc. [15], are the modern generation threats for CPPS. According to the security experts, we are heading towards a *Cyber Pearl Harbor* [151]. So the advancement of the cyber weapons are now considered major threats for the security and resiliency of the SG.

The entities in the power system environment are divided into two groups, i.e., the terrorists and the authorities. The terrorists are the entities with heinous intention to cause damages in the CPPS. The other entity is expected to protect the CPPS from these damages or to reduce these damages by protecting significant components. The complex interaction between these entities can be represented using game theory. Game theory can help the power system operators to understand the attackers' motives, moves, and thus they can strengthen the security of the CPPS. A lot of research works are going on for improving the stability, security, and resiliency of smart grid under different uncertain conditions [100]–[104]. Different forms of games are formulated to solve the challenges in the CPPS. One-shot attack [113], multi-stage attack [105], simultaneous attack [114], sequential attack [115], malicious false data injection attack [116] and coordinated attack are different types of attack schemes for applying in the power grid. Proper solution to the games adopting these attack schemes are important for the authorities to increase the protection and avoid the damages. To solve different game theoretic problems, different approaches are used such as, neural networks, deep neural networks, approximate dynamic programming, adaptive dynamic programming, and reinforcement learning [22]–[29]. So, the use of game theory can aid the power system operators and planners to resolve the challenging issues of

the advanced and complex electric power grid and increase the security and resiliency.

Recently, several research efforts [46], [64], [105]–[112] have been made based on Markov decision processes (MDPs), and game theory to make the security of the CPPS stronger against malicious attacks. In [106], authors used static game theoretic solution concept, called *Shapley value*. It represents coalition formation game theory in assessing the components' criticality for system's overall reliability and further maintenance focuses. In [64], [107], [108], authors implemented stochastic games for power grid protection against malicious attacker which are two-player zero-sum games and one-shot processes. Q-learning is applied to solve these games with optimal policies for the attacker. In [105], authors studied a multistage game between the adversaries that uses unique utility function to find the optimal attack sequences from the attacker's perspective. In this work, the defender's action is passive and predefined throughout the learning process of the attacker. In [109], authors modeled a strategic honeypot game for distributed DoS (denial of service) attacks in the SG. They analyzed the interactions between the attacker and the defender in the SG communication network. In [110], authors used iterated/repeated game (one-shot process) to analyze cyber-switching attacks and mitigation in SG systems based on zero-determinant strategies. In [46], [111], authors used reinforcement learning for sequential attack against power grid networks where only the attacker's action is considered. In [112], authors proposed a single stage game which is a zero-sum game in nature for risk modeling and mitigation for the SG security. In this work, authors also mentioned the multistage repeated game and Bayesian game as potential extensions of the work. So, MDPs and game theories are increasingly used in the SG security to conduct the interdisciplinary research between the machine learning and power system. In cyber-physical power system security,

very few papers have reported two active players in the repeated games. Those games do not consider costs of actions, budgets, and strengths as the factors for the pay-off evaluation. The impact of the attack policies on the power system has also been neglected, but is somehow very important to the physical system itself.

Based on the aforementioned literature, we propose a repeated game between the adversaries and use reinforcement learning-based solution to overcome the weaknesses. The main contributions of this paper are summarized below:

1. We proposed a repeated game model incorporating the cost of attack and defense actions and the average loss due to the attack in the system. The consideration of the attack and the defense costs will provide clear insight about the resource allocation and consumption for the attack and the defense schemes in the power system. This game is also applicable for the players with limited access to the information of the opponents. We incorporated budget constraints in the game model so that the players could cope with limited resources for a more realistic power system game theoretic environment. The idea of error and missing detection in the power system is introduced in the game model and demonstrated in the implementation. These detections will help the system operators evaluate the performance of the protection scheme. A comparative study between the overall game payoff and an existing approach is provided which will help the power system operators choose their mixed strategy actions.
2. We implement the interactions between the adversaries by formulating a two-person repeated game in the SG. The main advantage of using a repeated game is that it

does not require any information except what is shared between them. Unlike most of the existing literature, the repeated game conducts multiple repetitions instead of one-shot game, and uses individual utility functions to evaluate the payoffs. Then we design the solution of the game using a reinforcement learning algorithm which helps to achieve the outcome in favor of the adversaries. Favorable outcome means achieving higher payoff compared to the other player. The policies that result in favorable outcome to the players are significantly important for proposing a protection scheme and identifying the critical elements. Both players participate actively in the game and take action independently. They adopt a sequential action scheme unlike the existing literature and learn the policies based on the repeated game process. We also provide alternative action choices for the adversaries for different attack and defense strengths.

3. Realistic and crucial factors, like costs, allocated budgets, and strengths of the adversaries, have been neglected in the existing literature. In our work, costs are parameterized as a certain percentage of total loading capacity of the system. Budgets are parameterized as a number of limited actions. Strengths of the players are parameterized as a value ranging from zero to one. Then these parameters are incorporated as constraints to calculate the mixed strategy payoffs of the adversaries.
4. After analyzing the interactions and learning the policy, we illustrate the impact of the learned attacker's policy in the power system. Very few existing works reported the post attack effects (e.g., voltage, current, and power) based on the learned policies in the adversarial game. We simulate the learned attack policies in the PowerWorld

simulator. For evaluation of the impact, voltage violation is considered as the evaluation metric. This illustration provides insight about the loss caused by the cyber attack on the power system.

The rest of the paper is organized as follows. Section 4.2 explains the theoretical problem formulation of repeated game framework which involves strategic analysis of repeated game and its solution based on reinforcement learning. In Section ??, the results from the simulation case studies are explained for both strategic analysis of the game and reinforcement learning based solution of the game. Finally, Section ??, concludes the article with discussions and future works.

## 4.2 Problem formulation and implementation

### 4.2.0.1 Strategic analysis of repeated game behavior in smart grid security

The modeling of the game is adopted from [152]. In order to formulate the repeated game in power system, the game payoff function can be given as:

$$\Pi_i^t = \{C_m^t, C_i^t, C_a^t, C_w^t, P_a, U_a, U_i\} \in \{S_A^\tau, S_I^\tau\} \quad (4.1)$$

The parameters detail is given in Table 4.1.

Table 4.1. Parameter details for attacker-defender repeated game

Parameter	Detail
$C_m^t$	Cost of starting the attack detection system
$C_i^t$	Average generation loss when transmission line $i$ is attacked
$C_a^t$	The cost of attackers for attacking
$C_w^t$	The cost of attackers for not attacking
$U_a$	Attacking transmission line's payoff
$U_i$	Defending transmission line's payoff
$P_a$	Attacker's payoff
$S_A^\tau$	Attacker's actions space
$S_I^\tau$	Defender's actions space

To decide whether to attack or not, the attacking agent needs to satisfy two specific conditions. The first condition is given below:

$$\sum_{t=1}^T \sum_i C_w^t < P_a - \sum_{t=1}^T \sum_i C_a^t \quad (4.2)$$

where,  $P_a = \sum_{t=1}^T \sum_i C_i^T$ . The second condition that the attacking agent needs to satisfy is given as:

$$\sum_{t=1}^T \sum_i C_i^t \gg \sum_{t=1}^T \sum_i C_m^t \quad (4.3)$$

The payoff matrix for different strategies of the attacker and the defender can be shown as follows:

$$\begin{bmatrix} \sum_{t=1}^T \sum_i (P_a - C_a^t, U_i - C_i^t) & \sum_{t=1}^T \sum_i (-U_a, U_i - C_m^t) \\ \sum_{t=1}^T \sum_i (C_w^t, U_i) & \sum_{t=1}^T \sum_i (C_w^t, U_i - C_m^t) \end{bmatrix} \quad (4.4)$$

Here the columns and the rows in the payoff matrix represent the defender's and the attacker's possible action strategies respectively. In each matrix, the first and the second entity represents the attacker's and the defender's payoffs respectively.

**Nash equilibrium strategies in game model** In the game model, the attacker will achieve the maximum payoffs if the defenders do not start the attack detection system (ADS). Moreover, it is not advised for the defenders to start the ADS for a long time due to resource consumption. The defender's payoff will be higher if the transmission lines do not start the ADS. The solution of the payoff matrix in (4.4) will be found in the NE point. The game model has no pure strategy in NE. For pure strategy Nash equilibrium the game players should satisfy the following conditions:

$$P_a - \sum_{t=1}^T \sum_i C_a^t > \sum_{t=1}^T \sum_i C_w^t \quad (4.5)$$

and

$$\sum_{t=1}^T \sum_i -U_a < \sum_{t=1}^T \sum_i C_w^t \quad (4.6)$$

$$\sum_{t=1}^T \sum_i U_i - C_a^t < \sum_{t=1}^T \sum_i U_i - C_m^t \quad (4.7)$$

$$\sum_{t=1}^T \sum_i U_i > \sum_{t=1}^T \sum_i U_i - C_m^t \quad (4.8)$$

According to equation 4.5 and 4.6, attackers does not have any dominant strategies. For the defender, according to equation 4.7 and 4.8, defenders does not have any dominant strategies as well. Let us define the probability of attacking with  $\sigma_i$ , probability of starting the ADS at transmission line  $i$  with  $\delta_i$ , probability of not starting the ADS with  $(1 - \delta_i)$  and probability of not attacking with  $(1 - \sigma_i)$ . So, the mixed strategy payoffs are equivalent to the pure strategy payoffs according to the weighted average probability. The attacker's and the defender's mixed strategy payoffs are given by:

$$U_A = \sum_{t=1}^T \sum_i (P_a - C_a^t)(1 - \delta_i)\sigma_i - U_a\delta_i\sigma_i + C_w^t(1 - \sigma_i) \quad (4.9)$$

and

$$U_I = \sum_{t=1}^T \sum_i C_i^t\delta_i\sigma_i + U_i - C_m^t\delta_i - C_i^t\sigma_i \quad (4.10)$$

Now mixed strategy Nash equilibrium exists for the game model. To solve the mixed strategy Nash equilibrium extreme value method is used:

$$\frac{\partial U_A}{\partial \sigma_i} = \sum_{t=1}^T \sum_i (P_a - C_a^t)(1 - \sigma_i) - U_a\delta_i - C_w^t = 0 \quad (4.11)$$

$$\frac{\partial U_I}{\partial \delta_i} = \sum_{t=1}^T \sum_i C_i^t\sigma_i - C_m^t = 0 \quad (4.12)$$

**Attack and defense budget** To limit the attack and the defense actions, the attacker and the defender are assigned a certain budget. Let us consider the budget is a certain percentage of the total generation capacity of the power system. The attacker's budget and the defender's budget are defined by  $C_B$  and  $C_D$  respectively. The attack and the defense constraints are given as follows:

$$\sum_{t=1}^T \sum_i (C_a^t + C_w^t) \leq C_B \quad (4.13)$$

Where,  $C_a^t$ ,  $C_w^t$ , and  $C_B$  are the cost of attacking, cost of not attacking, and attacker's total budget respectively. Similarly, for the defender, the defense constraint is given as follows:

$$\sum_{t=1}^T \sum_i C_m^t \leq C_D \quad (4.14)$$

Here,  $C_m^t$  and  $C_D$  are the cost of starting the ADS and total defender's budget respectively. After satisfying the attack condition and budget constraints, the attacker and the defender will choose the probabilities of attack and defense. The probabilities of attack and defense actions execution are determined by comparing with a random number in each repetition of the game. The decision of attack and defense is determined by the equations below:

$$\frac{C_B - \sum_{t=1}^T \sum_i (C_a^t + C_w^t)}{C_B} \leq \text{Rand1} \quad (4.15)$$

and,

$$\frac{C_D - \sum_{t=1}^T \sum_i (C_m^t)}{C_D} \leq \text{Rand2} \quad (4.16)$$

Here,  $\text{Rand1}$  and  $\text{Rand2}$  are the two different random numbers generated in every repetition to compare with the ratio of budget spent and the original budget of the players.

**Game model analysis: Payoffs** If there are  $\lambda$  transmission lines in the system, then the payoff of the game model is:

$$U_{\text{game}} = \sum_{t=1}^T \sum_{i=1}^{\lambda} (C_i^t \delta_i \sigma_i + U_i - C_m^t \delta_i - C_i^t \delta_i - C_i^t \sigma_i) \quad (4.17)$$

And, the payoff for the AM model is given as:

$$U_{all} = \sum_{t=1}^T \sum_{i=1}^{\lambda} (U_i - C_m^t) \quad (4.18)$$

To compare the performance of the game model with that of the AM model, we can define the increased extra payoff by subtracting  $U_{all}$  from  $U_{game}$  as represented by equation (4.19).

$$\begin{aligned} U &= U_{game} - U_{all} \\ &= \sum_{t=1}^T \sum_{i=1}^{\lambda} (C_i^t \delta_i \sigma_i + U_i - C_m^t \delta_i - C_i^t \sigma_i) - \sum_{t=1}^T \sum_{i=1}^{\lambda} (U_i - C_m^t) \\ &= \sum_{t=1}^T \sum_{i=1}^{\lambda} [(C_i^t \delta_i \sigma_i + U_i - C_m^t \delta_i - C_i^t \sigma_i) - (U_i - C_m^t)] \\ &= \sum_{t=1}^T \sum_{i=1}^{\lambda} [(C_i^t \delta_i \sigma_i + C_m^t (1 - \delta_i) - C_i^t \sigma_i)] \\ &= \sum_{t=1}^T \sum_{i=1}^{\lambda} [(1 - \delta_i)(C_m^t - C_i^t \sigma_i)] \end{aligned} \quad (4.19)$$

From equation (4.19) we can conclude that, if  $\delta_i = 1$ , then  $U = 0$ . It means, when all the transmission lines are under the protection, the whole network will obtain no extra payoffs. If the frequency of the attacker is higher than the defender, then it will lead the power system achieving a higher level of protection undoubtedly. If the transmission lines start the ADS frequently, the system will consume many resources for defense. In that case, both the attacker and the defender will get no extra payoffs. So, as a rational defender, an exceptionally strong protection system cannot bring extra payoffs to the power system. In the mean time, as a rational attacker, a higher frequency attack fails to win over long term.

**Depicting game model payoff and AM model payoff** Here we will compare the outcome of the game model with the AM model. The payoff is representing how much power the defender is saving from being attacked. Or how the defense model is reducing the power consumption from the system's total generation compared to the AM model. Game

model payoff represents the total savings of the entire power system in terms of generation power. On the other hand the AM model monitors all the transmission lines with ADS. In this regard, if  $U_{game} > U_{all}$ , then the game model payoff is saving more generation power than the AM model. If  $U_{all} > U_{game}$ , then the game model is saving less generation power than that with the AM model. If  $U_{game} = U_{all}$ , then both models are saving same amount of generation power.

**Detection** When the attacker attacks and the defender starts the ADS or vice-versa, it is considered as a correct detection. When the attacker attacks but the defender does not start the ADS, it is considered as a missing detection. Similarly, when the attacker does not attack but the defender starts the ADS it is considered as an error detection. These four conditions of detections can be represented by a  $2 \times 2$  matrix below:

Table 4.2. Different detections in the attacker-defender repeated game in smart power grid security

	Defender defends	Defender does not defend
Attacker attacks	Correct detection	Missing detection
Attacker does not attack	Error detection	Correct detection

#### 4.2.0.2 Solution of adversarial repeated game using machine learning

**Background of repeated game** *Repeated games*, refers a situation in which the same stage game (strategic form game or one-shot game) is repeated for some duration. These type of games are also called “*supergames*”. In other words, a *repeated game* is a set of multiple *one-shot games*. It is modeled to evaluate the logic of long-term interactions between the players. The basic idea of this game is that a player will take into account the effect of his/her current behavior on the opponent players’ future behavior. Any history, ex-

cept what is shared between the two players, can be disregarded. In CPPS, the adversaries (the attacker or the defender) do not need to know the full history of the opponent (limited access to the information). So the game can still reach the optimality (Nash equilibrium (NE)) with limited information about the opponent. Also, given the repetition of the one-shot game, the environmental information is updated from the previous repetition. So, no information is missing about the attacker-defender interactions from the previous repetitions. In [153]–[157], Repeated game is used in different areas of study. The repeated game can be expressed as

$$\Pi_i^T = \{C_m^T, C_i^T, C_a^T, P_a, U_a, U_d\} \in \{S_A, S_D\} \quad (4.20)$$

To solve a repeated game using multiagent reinforcement learning (RL), several algorithms can be used.

Table 4.3. Different RL algorithms used for different environment types in the existing multiagent RL related works.

Environment type	Game type	RL algorithm
Adversarial	Zero-sum	Minimax-Q [158], [159]
Adversarial	General-sum	Nash-Q or FFQ [160], [161]
Collaborative	General-sum	Joint action learner (JAL)[162]
Adversarial collaborative	General-sum	Joint action learner (JAL)[162]

Table 4.3, is showing different algorithms used for solving different multiagent RL environments. From (4.20), we get the repeated game function  $\Pi_i^T$ . The game model is adopted from [157]. The payoff matrix of the repeated game can be formulated as:

$$\begin{bmatrix} \sum_{t=1}^T \sum_i (P_a - C_a^t, U_i - C_i^t) & \sum_{t=1}^T \sum_i (-U_a, U_i - C_m^t) \\ \sum_{t=1}^T \sum_i (C_w^t, U_i) & \sum_{t=1}^T \sum_i (C_w^t, U_i - C_m^t) \end{bmatrix} \quad (4.21)$$

To solve the game from this payoff matrix at (4.21), the attacker's and the defender's mixed strategy payoffs can be derived as:

$$\begin{aligned}
 U_A &= \sum_{t=1}^T \sum_i (P_a - C_a^t)(1 - \delta_i)\sigma_i + (-U_a)\delta_i\sigma_i + C_w^t(1 - \sigma_i) \\
 &= \sum_{t=1}^T \sum_i (P_a - C_a^t)(1 - \delta_i)\sigma_i - U_a\delta_i\sigma_i
 \end{aligned} \tag{4.22}$$

and

$$\begin{aligned}
 U_D &= \sum_{t=1}^T \sum_i (U_d - C_i^t)(1 - \delta_i)\sigma_i + U_i(1 - \delta_i)(1 - \sigma_i) \\
 &\quad + (U_i - C_m^t)\delta_i(1 - \sigma_i) = \sum_{t=1}^T \sum_i C_i^t\delta_i\sigma_i + U_d - C_m^t\delta_i - C_i^t\sigma_i
 \end{aligned} \tag{4.23}$$

Here  $\delta_i$  and  $(1 - \delta_i)$  are the probability of defending and not defending, and  $\sigma_i$  and  $(1 - \sigma_i)$  are the probability of attacking and not attacking.  $C_w$  is the cost of not attacking which is considered as zero. And, the total loss caused by the attack is defined as,

$$P_a = \sum_{t=1}^T \sum_i C_i^t \tag{4.24}$$

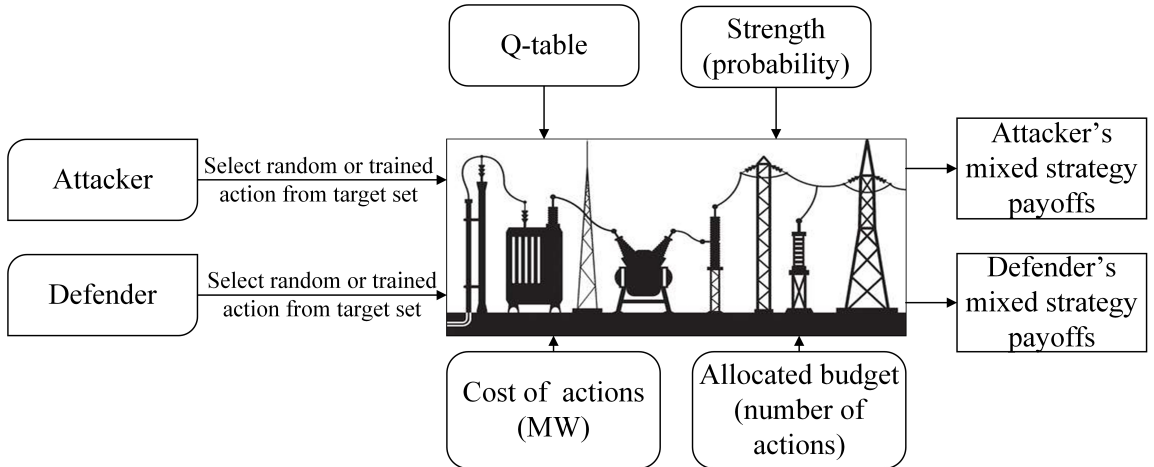


Figure 4.1. Attacker-defender interaction in a repetition of a repeated game.

**Proposed algorithm for the adversarial repeated game** The interaction between the adversaries in the power system in one repetition is shown in Figure 4.1. The adversaries take their actions from their associated target sets ( $S_A$  and  $S_D$ ). Their strength (probability of attacking and defending), cost of actions and allocated budgets' information is provided to calculate their mixed strategy payoffs. Figure 4.1 also visualizes the agent-environment interactions in the reinforcement learning framework. The attacker and the defender represents the agents, and the power system represents the environment. The mixed strategy payoffs are the feedback from the environment. Based on this feedback, the reward is assigned to the agents. The overall diagram of the gaming process using learning theory is presented in Figure 4.2.

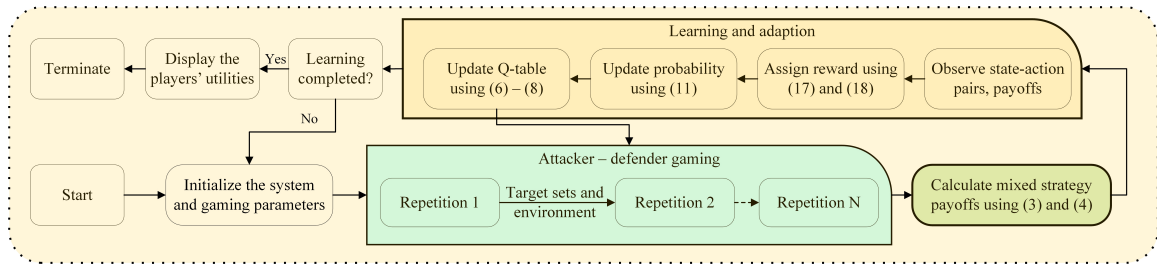


Figure 4.2. Overall block diagram of the reinforcement learning based solution of a repeated game in smart grid security.

Figure 4.2 is showing the process of solving a bi-level problem. In the upper level, the learning is helping to find the favorable outcome. In the lower level, gaming is conducted in between the adversaries. The main idea is to use reinforcement learning to learn the behavior of the adversaries for favorable outcome of the repeated game. The game showed in the lower block is repeated for transiting from one steady state to the next steady state. And at the end of repetition  $N$ , the game transits to the next run  $n = n + 1$ . The detailed interactions between the adversaries in one repetition are explained in Figure 4.1. The attacker's

target set, the defender's target sets, and the power system environment are updated from one repetition to another repetition. At the end of the repetitions, the history of the actions, payoffs, and rewards etc. from previous runs are updated for learning procedure.

**Proposed RL based solution** We formulate the repeated game as a zero-sum game and use Minimax-Q algorithm to solve it. Typical reinforcement learning is used to learn the optimal action strategies for favorable outcomes. The quality of the state for this game is given by

$$Q(s, a, d) = R + \gamma \sum_{s'} T(s, a, d, s') V_a(s') \quad (4.25)$$

$\gamma$  ranges from zero to one. The value of  $\gamma$  close to zero focuses on short term reward, and the value of  $\gamma$  close to one gives more emphasis on long term reward.  $T(s, a, d, s')$  is considered equal for all state transitions. The value of the state for the attacker can be calculated by

$$V_a(s') = \max_{\pi \in \Pi} \min_d \sum_a Q(s', a, d) \pi(a) \quad (4.26)$$

Similarly, the value of the state for the defender can be given by

$$V_d(s') = \min_{\pi \in \Pi} \max_a \sum_d Q(s', a, d) \pi(d) \quad (4.27)$$

Since this is a zero-sum game with weak duality,  $V_a(s') = V_d(s')$ . The problem in (4.26) can be solved using the similar value iteration algorithm as that in [126], [133],

$$\begin{aligned} & \max_{\pi} V_a(s') \\ & s.t. \sum_{a \in S_{N_a}} Q(s, a, d) \pi \geq V_a(s') \\ & \sum_{a \in S_{N_a}} \pi(a) = 1 \\ & \pi(a) \geq 0, \forall a \in S_{N_a} \end{aligned} \quad (4.28)$$

And the optimal policy is found by

$$\pi'(s) = \arg \max_a Q_{\pi}(s, a, d) \quad (4.29)$$

The probabilities of the state-action pairs are updated following the formula below

$$Pr(s, a, d) \leftarrow \frac{C(s, a, d)}{\sum_{a \in A, d \in D} C(s, a, d)} \quad (4.30)$$

The attacker's mixed strategy for a given state  $s$  will be:

$$\pi_A(s) = [Pr\{a(s) = a_1\}, \dots, Pr\{a(s) = a_N\}] \quad (4.31)$$

where

$$\sum_{i=1}^N Pr\{a(s) = a_i\} = 1 \quad (4.32)$$

Here,  $Pr\{a(s) = a_i\}$  is the probability of choosing attack action  $a_i$  in state  $s \in S_A$ , and  $\pi_A(s)$  represents the probability distribution over the attacker's action space associated with the state  $s$ . Similarly we can define the probability for the defender as well.

Algorithm 4 gives the steps for a repeated game in a run. The repetition continues till the resource expenditure by the players exceed the allocated budget or till there are any targets available in the target sets. In every repetition the adversaries interact and their associated utilities are calculated. Their actions are recorded and their probabilities are updated in every repetitions.

Algorithm 3 represents the Q-learning algorithm that is used to solve the repeated game. This Q-learning algorithm is learning the behavior of the players from their actions in the top layer of the gaming framework. In the bottom layer, the repeated game itself is being played multiple times. Algorithm 3 is initialized with the players' target sets as the input. The expected outputs of this algorithm are the optimal action sequences for the players along with their probabilities.

Then according to the exploration probability, actions are taken by the players either randomly or following the policy. The exploration probability is denoted by epsilon,  $\epsilon$ . The

---

**Algorithm 2: Adversarial repeated game in SG security**


---

**Input :** Test case, attacking probability, defending probability, budgets of the players, costs of attack and defense, attack and defense set

- 1 Initialize the Q-table, action counter, policies for each state-action pairs with uniform probability distribution;
- 2 **for** *The given system, and maximum number of run* **do**
- 3     Initialize the players' resource spent;
- 4     **for** *Maximum number of repetition and till the players' resource spent  $\leq$  Budget* **do**
- 5         Take actions from the agents' associated sets based on **Minimax-Q algorithm** from Algorithm (3);
- 6         Calculate the cost of actions using (4.38) ;
- 7         Update the resources spent from budget;
- 8         Calculate the utilities using (4.22) and (4.23);
- 9         Update the action sets from (4.34);
- 10        **if** *There are available resources for the attacker and the defender* **then**
- 11            Go to next repetition;
- 12        **else**
- 13            Exit the repetition loop and go to next run;
- 14        **end**
- 15     **end**
- 16 **end**

**Output:** The optimal actions with their associated probabilities;

---

final value of epsilon, ( $\epsilon_f$ ), is a very small positive number which ensures a very small exploration probability for the players to take very few random actions even at the end of the runs. With the execution of the actions, a transmission line is switched to out of service and another transmission line is defended from being attacked. After execution of the actions, the utilities for the players are calculated and their associated Q-values are updated. The probabilities for the associated state-action pair is also updated. This process is continued for the maximum number of runs.

**Design parameters** Design parameters of this repeated game solution based on reinforcement learning between the adversaries in the electric power grid are explained briefly in this subsection.

---

**Algorithm 3: Minimax-Q for repeated game**


---

**Input :** Attack and defense action sets

```

1 for Maximum number of run do
2   Initialize the attacker's and the defender's state;
3   for Maximum number of repetition do
4     if Prob = Explore then
5       Take random action from the available actions in the sets for both players;
6     else
7       Take attack action using (4.26);
8       Take defense action using (4.27);
9     end
10    Execute the action using (4.35);
11    Update action counter,  $C = C + 1$  and associated probabilities using (4.30);
12    Calculate the utilities;
13    Assign the reward based on (4.36);
14    Update the Q-value using (4.25);
15  end
16 end

```

**Output:** Output the optimal action policies for the attacker and the defender with their associated probabilities;

---

**Attack and defense sets** The attack and defense sets are the collection of targets for the attacker and the defender. Since, we use line switching attack (explained in subsection ??), transmission lines are considered as the target elements for attacking and defending. Every repetitions of the game is considered as the states of the game and represented by  $S$ , and

$$S = \{R_1, R_2, \dots, R_N\} \quad (4.33)$$

where,  $R_N$  is the repetition, and  $N$  is the number of total repetitions. The attackers and the defenders action spaces are represented by  $S_A$  and  $S_D$

$$\begin{aligned} S_A &= \{a_1, a_2, \dots, a_N\} \\ S_D &= \{d_1, d_2, \dots, d_N\} \end{aligned} \quad (4.34)$$

where,  $a_N$  and  $d_N$  are the attack and the defense actions in the  $N^{th}$  repetition, respectively.

**Reward** After each attack and defense action, reward is assigned by accessing the feedback of the actions from the environment.  $R_A(s, a, d)$ , represents the attacker's expected reward associated with the state  $s \in S$ , and attack and defense actions  $a \in S_A$ , and  $d \in S_D$ , respectively. Similarly,  $R_D(s, a, d)$  represents the defender's expected reward. The reward is designed based on the requirement. The states of this game are presented as a combination of the elements of the target sets. Whenever, an attack or defense action is triggered, the index of that transmission line in the target set switches to zero. The line status in a target set can be represented by  $s_t(l)$

$$s_t(l) = \begin{cases} X, & \text{if line } l \text{ is in-service at time } t \\ 0, & \text{if line } l \text{ is out-of-service at time } t \end{cases} \quad (4.35)$$

where  $X$  represents the transmission line number. In this zero-sum game, the reward is assigned opposite to each other. So, if the attacker's and the defender's mixed strategy payoffs are  $U_A$  and  $U_D$ , then the reward is assigned following the conditions below. When the desired outcome of the game is in favor of the attacker:

$$R^A(s, a, d) = \begin{cases} +1, & \text{if } U_A > U_D \\ 0, & \text{otherwise.} \end{cases} \quad (4.36)$$

and

$$R^D(s, a, d) = \begin{cases} -1, & \text{if } U_A < U_D \\ 0, & \text{otherwise.} \end{cases} \quad (4.37)$$

When the game is conducted expecting the outcome in favor of the defender, the reward assignment is just in the opposite way of (4.36) and (4.37).

**Allocated budget and the costs** The allocated budget for the attacker and the defender is defined by  $B_A$  and  $B_D$ . The budget is the amount of resource that limits the action space for the players. In this game, we considered limited number of actions according to the allocated budget for the players. For demonstration purpose we consider that the attacker and the defender can not take more than three actions. Cost is the amount of resource that the players consume while executing the actions. For simplicity, we consider the cost of the attack and defense action as a certain percentage of the total loading capacity of the system. In this game, if the cost of attack and defense action is defined by  $C_a$  and  $C_m$ ,

$$\begin{aligned} C_a &= 0.1 \times TLC \\ C_m &= 0.1 \times TLC \end{aligned} \tag{4.38}$$

where  $TLC$  represents the total loading capacity of the system. The attacking and defending probability is defined here by  $\sigma$  and  $\delta$ . These probabilities represent the strengths of the attacker and the defender. These strengths are used to calculate the mixed strategy payoffs of the players.

**Depicting the attacker's and the defender's mixed strategy payoffs, and Nash equilibrium** In Nash equilibrium, we compare the attacker's mixed strategy payoffs with the defender's mixed strategy payoffs. These payoffs represents how much power is saved due to the defense actions and how much power is lost due to the attack actions. Due to the rewarding condition  $R^D(a, d, s) = -R^A(a, d, s)$ , the proposed repeated game is a zero-sum game [163]. To solve a two-player stochastic game in normal form, one popular solution is closed-loop Nash equilibrium. Nash's theorem guarantees that the Nash equilibrium exists for static games. But for stochastic games, the possible number of strategies are infinite

[64], [133], [164]. For a reinforcement learning-based solution of a repeated game, the existence of the Nash equilibrium can be confirmed through the decision making process over time. In [165], it is showed that optimal defense strategy can be achieved adopting game-theoretic actor critic neural network for optimal defense and worst attack policy. There might be multiple optimal strategies for the repeated games in case of finite time horizon. While the attacker converges to its optimal action policies maximizing its payoffs, the defender converges to its optimal policies minimizing its payoffs. This confirms the existence of Nash equilibrium in the proposed game.

### 4.3 Simulation studies

#### 4.3.0.1 Strategic analysis of repeated game

The simulation is conducted using MATLAB R2018a on a standard PC with an Intel(R) Core(TM) i7-6700 CPU running at 3.40GHz and 24.0 GB RAM. *IEEE* 39-bus system is used as the simulation benchmark. This system contains 46 transmission lines. We divided the whole power system in two different sets. Both the attacker's set and the defender's set contain 23 transmission lines. Total generation capacity is defined as *TGC* which is 6150.05MW for *IEEE* 39-bus system. To define the costs of attacking, not attacking, and defending, we made the following assumptions:

- $C_a = C_w = C_m = 0.1\%$  of TGC
- Attack order is considered as 2. It means, two transmission lines will be attacked at the same time.

Simulation is conducted for different attack and defense probabilities and different attacker's and defender's budgets to observe the impact on the game model and the AM

model. The attack and the defense probabilities ranges from 0 to 1. And, the budget ranges for three different combinations, i.e.,  $C_B = C_D$ ,  $C_B > C_D$ , and  $C_B < C_D$ . Here  $C_B$  and  $C_D$  are representing the attacker's and the defender's budgets respectively.

#### 4.3.0.2 Simulation setup

All the possible combinations for the players' budgets and the probabilities of their actions are given in the table below:

Table 4.4. Possible combinations of the players' budget and the probabilities of their actions for conducting experiments in the repeated game

Number	Budget	Probability
1	$C_B = C_D$	$\delta_i = \sigma_i$
2	$C_B = C_D$	$\delta_i > \sigma_i$
3	$C_B = C_D$	$\delta_i < \sigma_i$
4	$C_B > C_D$	$\delta_i = \sigma_i$
5	$C_B > C_D$	$\delta_i > \sigma_i$
6	$C_B > C_D$	$\delta_i < \sigma_i$
7	$C_B < C_D$	$\delta_i = \sigma_i$
8	$C_B < C_D$	$\delta_i > \sigma_i$
9	$C_B < C_D$	$\delta_i < \sigma_i$

In order to test the characteristics of the attacker and the defender, we divide the whole experiments in two phases. In the first phase, the attacker and the defender select their possible actions space randomly in every repetition of the game. In the second phase, we randomly divide the test power system in two (the attacker's and the defender's) action spaces and then conduct the experiments. For both phases, we vary the attacking and the defending probabilities from 0 to 1 and compare the game model payoffs and the AM model payoffs. For random attack and defense set with equal attacker's and defender's budgets and equal attacking and defending probabilities, the observed results are given in Table 4.5.

Table 4.5. Observation of the game model payoff and the AM model payoff for equal budget and equal attacking and defending probabilities. The attacker's and the defender's budgets for this case studies are  $0.1 \times TGC$

Case study	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observations
1.01	0.1	0.1	$U_{game} > U_{all}$
1.02	0.2	0.2	
1.03	0.3	0.3	
1.04	0.4	0.4	
1.05	0.5	0.5	
1.06	0.6	0.6	
1.07	0.7	0.7	
1.08	0.8	0.8	
1.09	0.9	0.9	
1.10	1	1	$U_{game} = U_{all}$

Now, let us take a look at the characteristics of the payoffs and the budget during the repetitions of the game for the case study 1.01 from Table 4.5. Attack flags are measured by equations (4.2) and (4.3). Detection is measured according to the assumptions of Table 4.2.

**Simulation results** The simulation is conducted for all the cases from Table 4.5 and for different attacker's and defender's budgets. Figure 4.3(a) from shows the attack flag due to the initiated attacks. The first two attacks were successful in satisfying the conditions of (4.2) and (4.3). Figure 4.3(b) shows the residual generation after the attack and the defense actions. Figure 4.4(a) shows the players' mixed strategy payoffs for individual repetitions. Figure 4.4(b) gives the players' winning percentage for the whole game. Figure 4.5(a) and 4.5(b) show the attacker's and the defender's expenditure of the resources from the allocated budget. In Figure 4.5(b), from 1<sup>st</sup> to 15<sup>th</sup> repetitions, the defender's budget spent was stable and did not increase because the defender did not perform defense action in

those repetitions. Figure 4.6(a) provides the comparison between the game model payoffs and the AM model payoffs for individual repetitions. Though in the beginning both the models had equal payoffs, at the end the game model has higher payoffs than that of the AM model. Figure 4.6(b) provides the percentages of different detections of the attack detection systems.

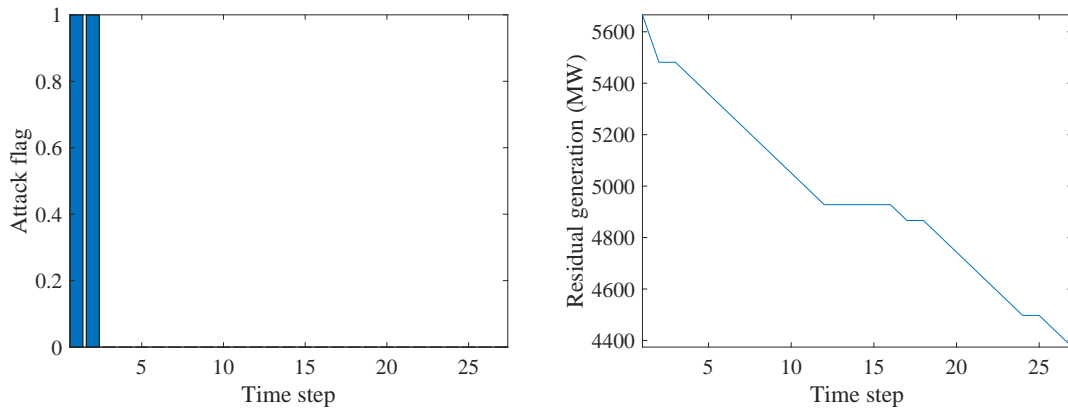


Figure 4.3. (a) Attack flag for the repeated game. (b) Residual generation after attack-defense.

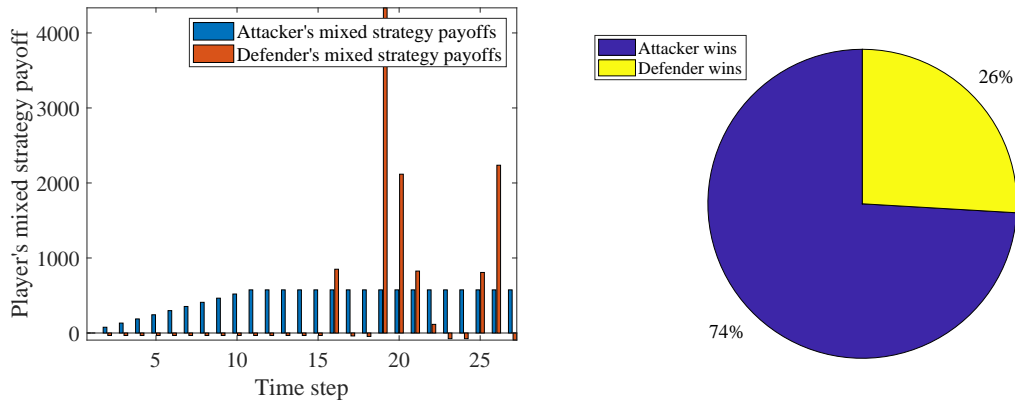


Figure 4.4. (a) Players' mixed strategy payoff for individual repetitions and (b) winning percentage from the mixed strategy payoffs. The attacker's and the defender's budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = \delta_i = 0.1$ .

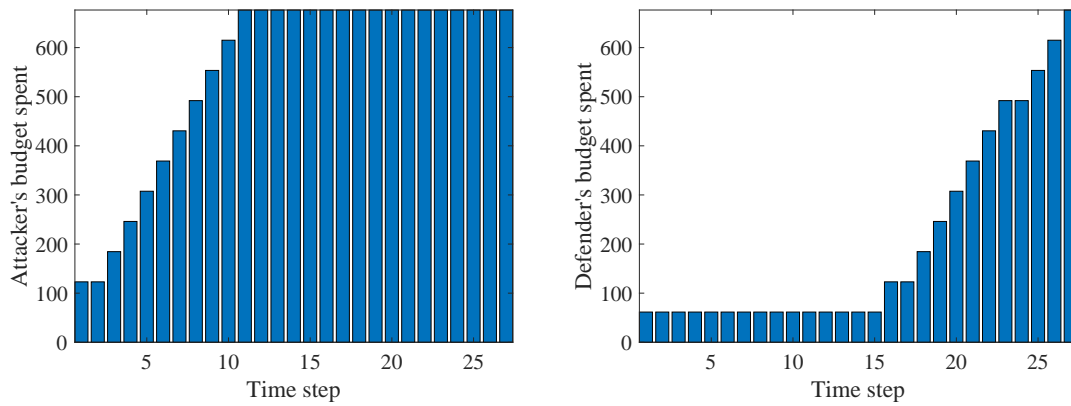


Figure 4.5. (a) The attacker's budget and (b) the defender's budget spent with each repetitions. The attacker's and the defender's budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = \delta_i = 0.1$ .

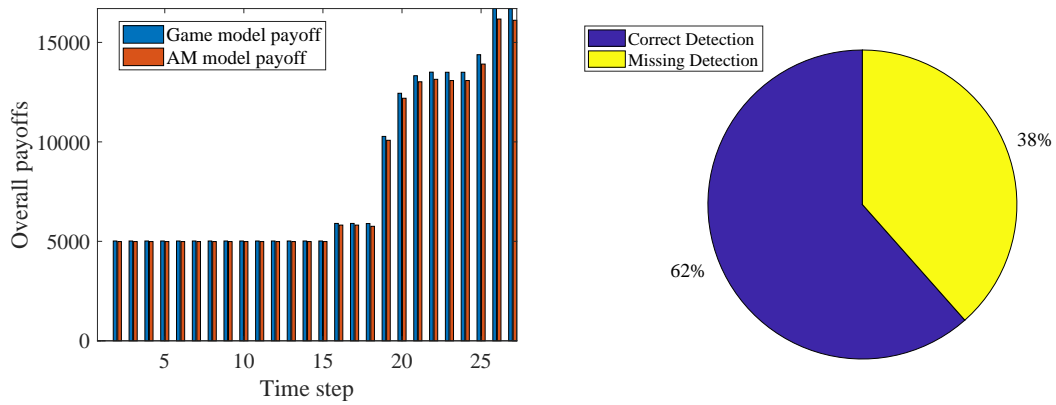


Figure 4.6. (a) Overall payoff of the game model and AM model with each repetitions. (b) percentage of correct, error and missing detections in the game. The attacker's and the defender's budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = \delta_i = 0.1$ .

Table 4.6. Observation of game model payoff and AM model payoff for equal budget and different attacking and defending probability

Case study	Attacker's budget, $C_B$	Defender's budget, $C_D$	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observation
2.01	$0.1 \times TGC$		0.1	1	$U_{game} < U_{all}$
2.02			0.2	0.9	$U_{game} < U_{all}$
2.03			0.3	0.8	$U_{game} < U_{all}$
2.04			0.4	0.7	$U_{game} < U_{all}$
2.05			0.5	0.6	$U_{game} < U_{all}$
2.06			0.6	0.5	$U_{game} > U_{all}$
2.07			0.7	0.4	$U_{game} > U_{all}$
2.08			0.8	0.3	$U_{game} > U_{all}$
2.09			0.9	0.2	$U_{game} > U_{all}$
2.10			1	0.1	$U_{game} = U_{all}$

Now let us take a look at the characteristics of the payoffs and budget during the repetitions of the game for the case study 2.10 from table 4.6. Attack flags are measured by equation 4.2 and 4.3. Detections are measured according to the assumptions of table 4.2.

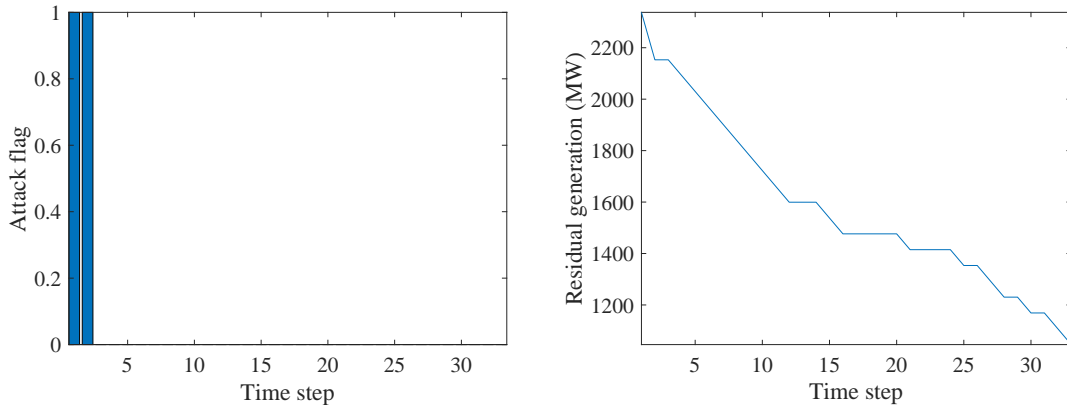


Figure 4.7. (a) Attack flag and (b) residual generation after attack-defense. The attackers' and defenders' budget  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively.

Figure 4.7(a) represents the attack flag for each time step that satisfies the two conditions for attack. Figure 4.7(b) represents the residual generation with each time step. We

can notice even if the attack flag is zero after the second time step, the generation is reducing. This is because, the players are consuming resources in terms of generation from the system.

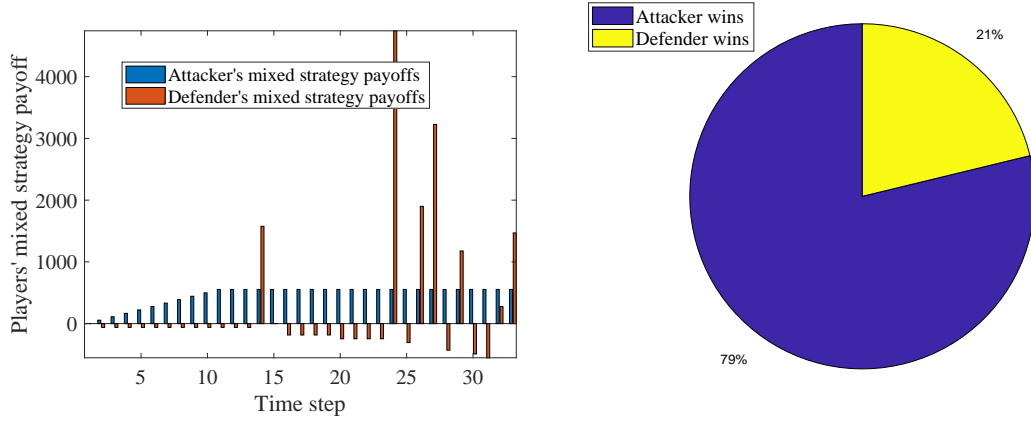


Figure 4.8. (a) Players' mixed strategy payoff for individual repetitions and (b) winning percentage from the mixed strategy payoffs. The attackers' and defenders' budget was  $C_B = C_D = 0.1\%$  of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively.

Figure 4.8(a) represents the players mixed strategy payoffs in every time step. From this figure we can see defender's mixed strategy payoffs in some of the time steps are negative. This mixed strategy payoffs are calculated using equation 4.9 and 4.10. The negative mixed strategy payoffs states that, the cost of starting the attack detection system and average generation loss when transmission line is attacked is very high than the defending transmission line's payoff. Figure 4.8(b) represents the players winning percentage in the mixed strategy payoffs.

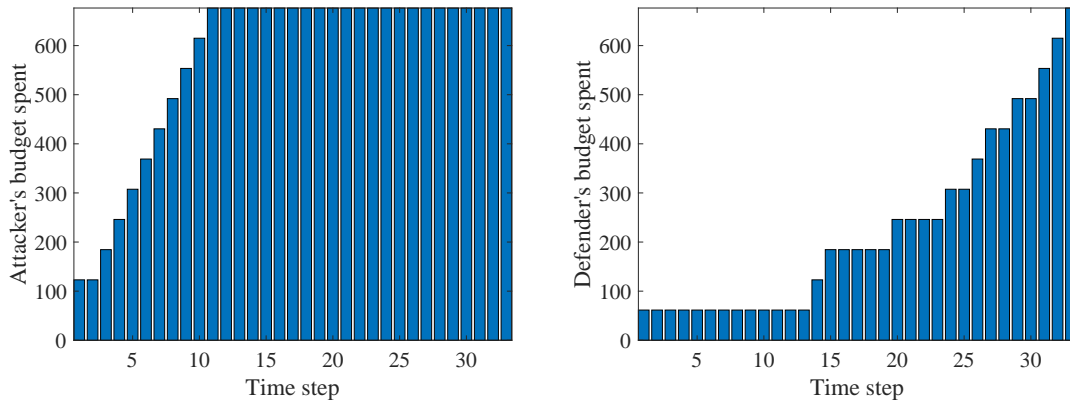


Figure 4.9. (a) Attackers' and (b) defenders' budget spent with each repetitions. The attackers' and defenders' budget was  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively.

Figure 4.9(a) and 4.10(a) represents the attacker's and the defender's budget spent for each time step. For both players the budget spent is saturated once they finished absorbing the allocated budget for them.

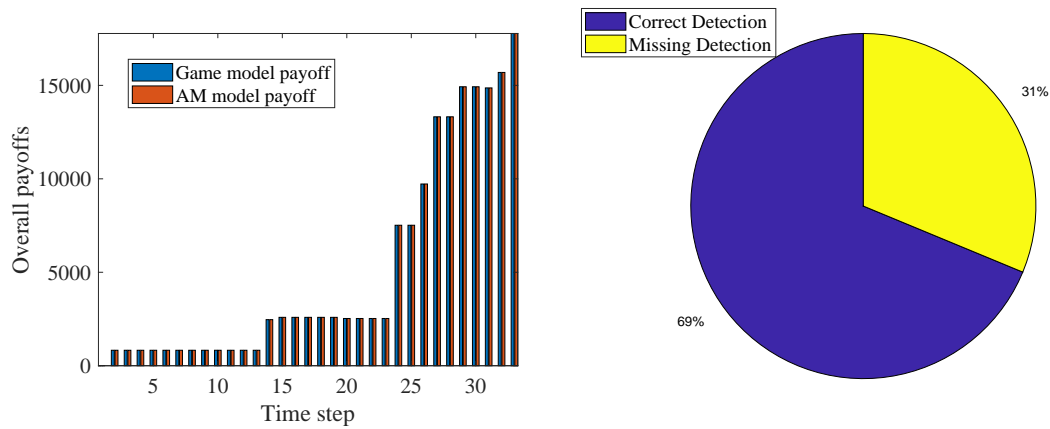


Figure 4.10. (a) Overall payoff of the game model and AM model with each repetitions and (b) percentage of correct, error and missing detections in the game. The attackers' and defenders' budget was  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability was  $\sigma_i = 0.1$  &  $\delta_i = 1$  respectively.

Figure 4.10(a) represents the overall payoff of the game model and the AM model with each repetition. We can see for given condition (players' budget and strengths) the AM

model and the game model achieves equal payoffs. This represents that in this condition both model performs equally. Figure 4.10(b) represents the error, correct and missing detections. From the figure, we can see that there is no error detection, but correct and missing detection.

Now let us take a look at the characteristics of the payoffs and budget during the repetitions of the game for the case study 2.05 from table 4.6. Attack flags are measured by equation 4.2 and 4.3. Detection are measured according to the assumptions of table 4.2.

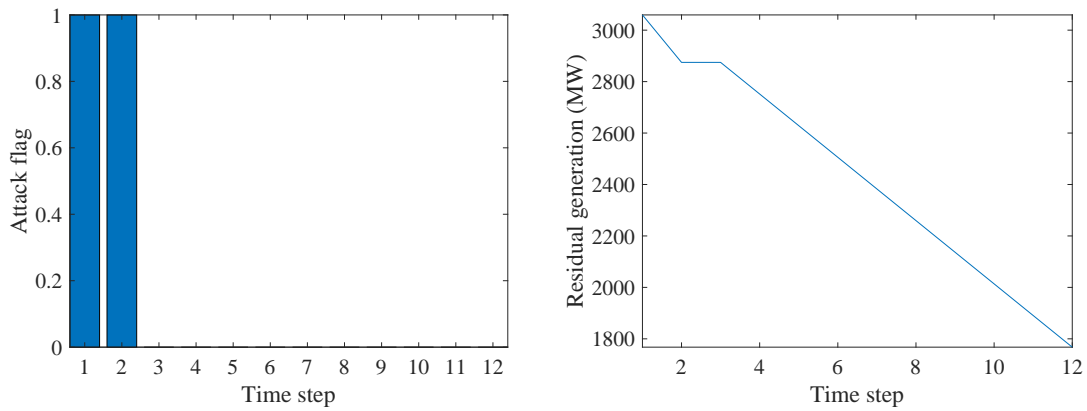


Figure 4.11. (a) Attack flag and (b) residual generation after attack-defense. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ .

Figure 4.11(a) represents the attack flag and figure 4.11(b) represents the residual generation with each repetitions. Figure 4.12(a) represents the mixed strategy payoffs for the attacker and the defender. From the mixed strategy payoffs and figure 4.12(b) we can see that, the attacker's mixed strategy payoff is higher than the defender's mixed strategy payoffs in each repetitions.

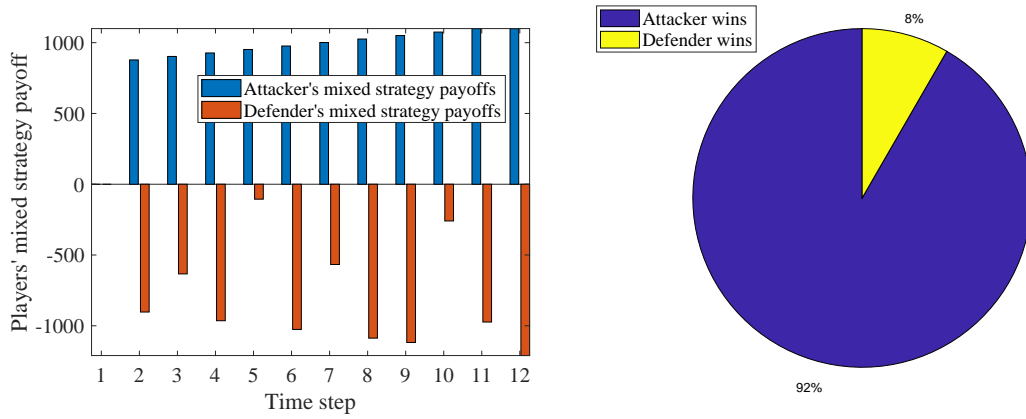


Figure 4.12. (a) Players' mixed strategy payoff and (b) winning percentage from the mixed strategy payoffs. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ .

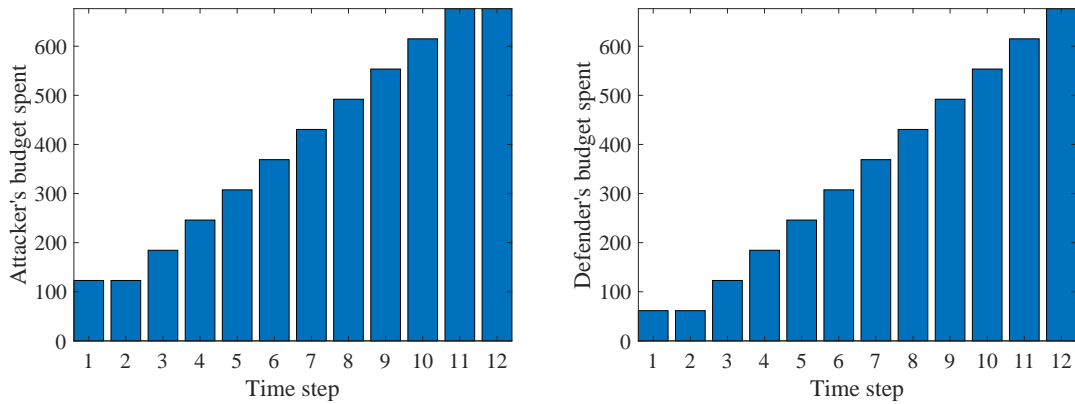


Figure 4.13. (a) Attackers' and (b) defenders' budget spent with each repetitions. The attackers' and defenders' budget is  $C_B = C_D = 0.1$  % of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ .

Figure 4.13(a) and 4.13(b) represents the attackers and the defender's budget spent in each repetitions.

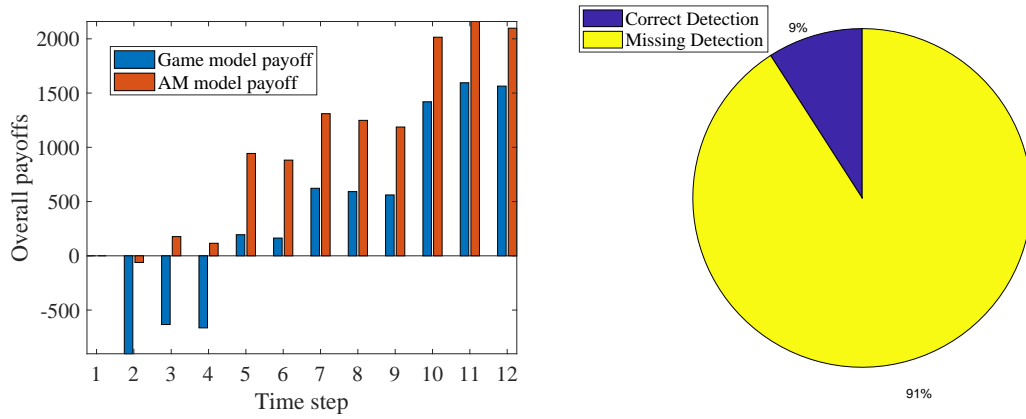


Figure 4.14. (a) Overall payoff of the game model and AM model with each repetitions and (b) percentage of correct, error and missing detections in the game. The attackers' and defenders' budget is  $C_B = C_D = 0.1\%$  of total generation capacity. The attacking and the defending probability is  $\sigma_i = 0.6$  &  $\delta_i = 0.5$ .

Figure 4.14(a) represents the overall payoff comparison between the game model and the all monitoring model. From the figure we can conclude that at the end of the repetitions, the AM model payoff is higher than the game model payoff. It means, the AM model saves higher amount of generation power than the proposed and implemented game model. Figure 4.14(b) represents the detection rates for the correct detections, missing detections, and error detections. Table 4.7 shows the observations for comparing the game model payoffs and AM model payoffs. The attacker's budget is higher than the defender's budget. Both attacker's and defender's strengths are equal. It is observed that, the game model payoff is higher than the AM model payoff with this game settings, except when both agents' strengths are 1, game model payoffs and AM model payoffs are equal.

Table 4.7. Observation of game model payoff and AM model payoff for different attacker's and defender's budget and equal attacking and defending probability

Case study	Attacker's budget, $C_B$	Defender's budget, $C_D$	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observation
3.01	$0.2 \times TGC$	$0.1 \times TGC$	0.1	0.1	$U_{game} > U_{all}$
3.02			0.2	0.2	$U_{game} > U_{all}$
3.03			0.3	0.3	$U_{game} > U_{all}$
3.04			0.4	0.4	$U_{game} > U_{all}$
3.05			0.5	0.5	$U_{game} > U_{all}$
3.06			0.6	0.6	$U_{game} > U_{all}$
3.07			0.7	0.7	$U_{game} > U_{all}$
3.08			0.8	0.8	$U_{game} > U_{all}$
3.09			0.9	0.9	$U_{game} > U_{all}$
3.10			1	1	$U_{game} = U_{all}$

Table 4.8. Observation of game model payoff and AM model payoff for different attacker's and defender's budget and different attacking and defending probability

Case study	Attacker's budget, $C_B$	Defender's budget, $C_D$	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observation
4.01	$0.2 \times TGC$	$0.1 \times TGC$	1	0.1	$U_{game} = U_{all}$
4.02			0.9	0.2	$U_{game} > U_{all}$
4.03			0.8	0.3	$U_{game} > U_{all}$
4.04			0.7	0.4	$U_{game} > U_{all}$
4.05			0.6	0.5	$U_{game} > U_{all}$
4.06			0.5	0.6	$U_{game} < U_{all}$
4.07			0.4	0.7	$U_{game} < U_{all}$
4.08			0.3	0.8	$U_{game} < U_{all}$
4.09			0.2	0.9	$U_{game} < U_{all}$
4.10			0.1	1	$U_{game} < U_{all}$

Table 4.8 represents the observations when the attacker's budget is higher than the defender's budget and, their strengths are varied for different case studies to observe the game model payoffs and AM model payoffs.

Table 4.9. Observation of game model payoff and AM model payoff for different attacker's and defender's budget and equal attacking and defending probability

Case study	Attacker's budget, $C_B$	Defender's budget, $C_D$	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observation
5.01	$0.1 \times TGC$	$0.2 \times TGC$	1	1	$U_{game} > U_{all}$
5.02			0.9	0.9	$U_{game} > U_{all}$
5.03			0.8	0.8	$U_{game} > U_{all}$
5.04			0.7	0.7	$U_{game} > U_{all}$
5.05			0.6	0.6	$U_{game} > U_{all}$
5.06			0.5	0.5	$U_{game} > U_{all}$
5.07			0.4	0.4	$U_{game} > U_{all}$
5.08			0.3	0.3	$U_{game} > U_{all}$
5.09			0.2	0.2	$U_{game} > U_{all}$
5.10			0.1	0.1	$U_{game} = U_{all}$

Table 4.9 shows the observations where defender's budget is higher than the attacker's budget and their strengths are equal.

Table 4.10. Observation of game model payoff and AM model payoff for different attacker's and defender's budget and different attacking and defending probability

Case study	Attacker's budget, $C_B$	Defender's budget, $C_D$	Defending probability, $\delta_i$	Attacking probability, $\sigma_i$	Observation
6.01	$0.1 \times TGC$	$0.2 \times TGC$	1	0.1	$U_{game} = U_{all}$
6.02			0.9	0.2	$U_{game} > U_{all}$
6.03			0.8	0.3	$U_{game} > U_{all}$
6.04			0.7	0.4	$U_{game} > U_{all}$
6.05			0.6	0.5	$U_{game} > U_{all}$
6.06			0.5	0.6	$U_{game} < U_{all}$
6.07			0.4	0.7	$U_{game} < U_{all}$
6.08			0.3	0.8	$U_{game} < U_{all}$
6.09			0.2	0.9	$U_{game} < U_{all}$
6.10			0.1	1	$U_{game} < U_{all}$

Table 4.10 shows the observations where the defender's budget is higher than the attacker's budget and their strengths are varied to observe the comparison between the game model payoffs and the AM model payoffs.

**Observations** For both random and fixed attack and defense set the following observations are found:

- For  $\delta_i = 1$ ,  $U_{game} = U_{all}$ ,  $U_{extra} = 0$ . Which means, the game model and the AM model save the same amount of generation power.
- For  $\delta_i > \sigma_i$ ,  $U_{game} > U_{all}$ ,  $U_{extra} > 0$ . Which means, the game model saves higher amount of generation power than the AM model.
- For  $\delta_i < \sigma_i$ ,  $U_{game} < U_{all}$ ,  $U_{extra} < 0$ . Which means, the game model saves less amount of generation power than the AM model.
- For  $\delta_i = \sigma_i$ ,  $U_{game} > U_{all}$ ,  $U_{extra} > 0$ . Which means, the game model saves higher amount of generation power than the AM model.

Now, we will compare the overall payoff for three different strategies for the game model payoff and the AM model payoff.

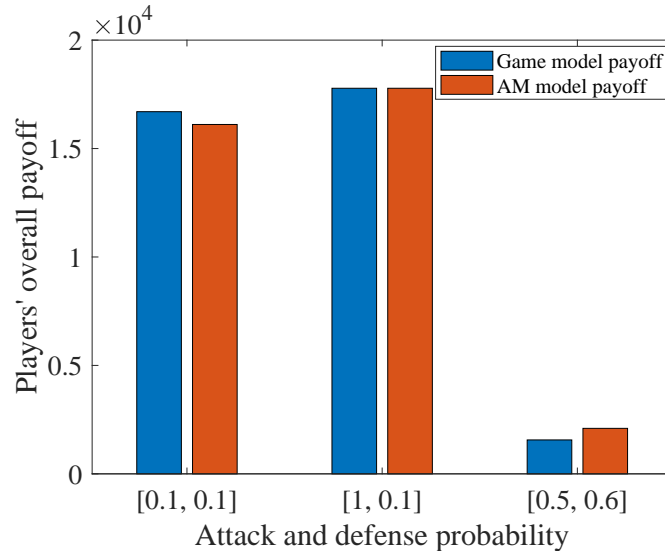


Figure 4.15. Comparison of the game model payoff and the AM model payoff for three different attack and defense strategies.

From Figure 4.15, we can see that, the game model payoff and the AM model payoffs change with the attack and defense probabilities. These relations between the game model and the AM model payoffs also validate the observations found in this subsection. For some of the strategies, the game theory model gives higher payoff than that of AM model payoff. The game model payoff is favorable to the defender for those strategies. Here the attacker's and the defender's budgets are equal. The attack and the defense probabilities are given in the X-axis of the figure.

#### 4.3.0.3 Q-learning based solution for a repeated game

The simulation is conducted using MATLAB R2018a on a standard PC with an Intel(R) i7-6700 CPU running at 3.40GHz and 24.0 GB RAM. There are mainly three loops in the simulation: rounds, runs, and repetitions. The repetitions loop ends when the allocated budget is fully consumed by the players. The runs loop is the main loop where the players (the attackers and the defenders) and the environment (power system) interacts to converge to the optimal policies. As the number of runs increases, the players tend to learn the optimal policy. At the end of the runs loop, the players reach to the Nash Equilibrium point. The runs loop is repeated in several rounds in the rounds loop to get different optimal policies. The number of repetitions also represent the number of actions that the attacker and the defender takes. The number of runs represents the number of trials required in the learning process.

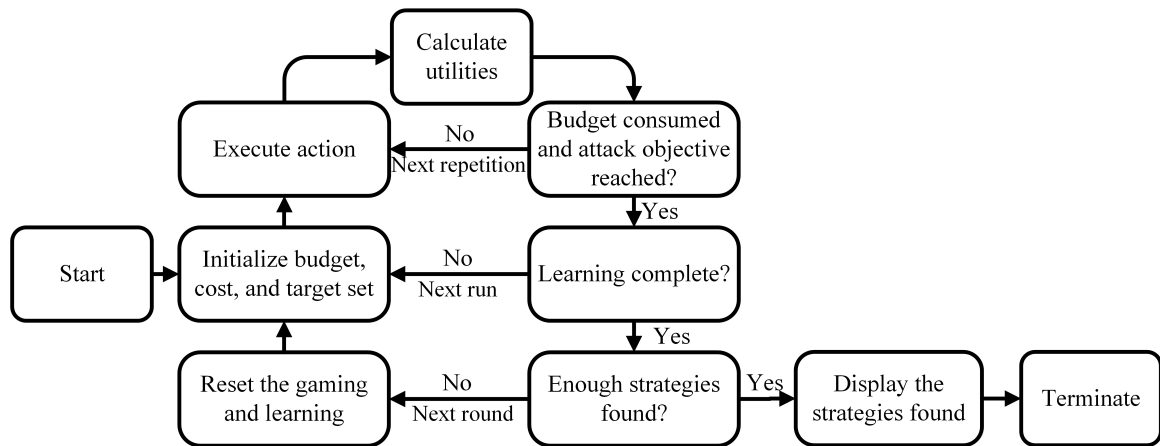


Figure 4.16. Simulation loops are explained in this figure. The simulation contains repetitions, runs, and round loops.

A Q-table is required in the learning process using Q-learning algorithm. Q-table is defined with the attacker's and the defender's state-action pairs, action counters, probabilities, and Q-values. The discount factor,  $\gamma$  is applicable for both of the players learning procedure. So, both of the players decide the value of  $\gamma$ . The attacker and the defender are allowed to take only one action in a repetition.

The outcome of this repeated game, solved by reinforcement learning, is analyzed mainly from two different perspectives. First, we analyze the outcome from the game theoretic viewpoint where, the attacker's and the defender's utility is compared. Next, we analyze the post-attack effects in the simulated power system. In this section, we analyze the outcome of the game from game theoretic viewpoint. We conduct different case studies under different conditions to find out the favorable outcome for the attacker and the defender. Figure 4.17, is showing the case studies and the conditions associated with the repeated game solution using reinforcement learning. Branches in the left side under the case studies are the conditions, where the outcome of the game is in favor of the attacker

( $U_A > U_D$ ). Branches in the right side are the conditions, where the outcome of the game is in favor of the defender ( $U_A < U_D$ ).  $\delta_i$  and  $\sigma_i$  represent the probability (strength) of defending and attacking, respectively. For different attacking and defending strengths, the game simulation is conducted.

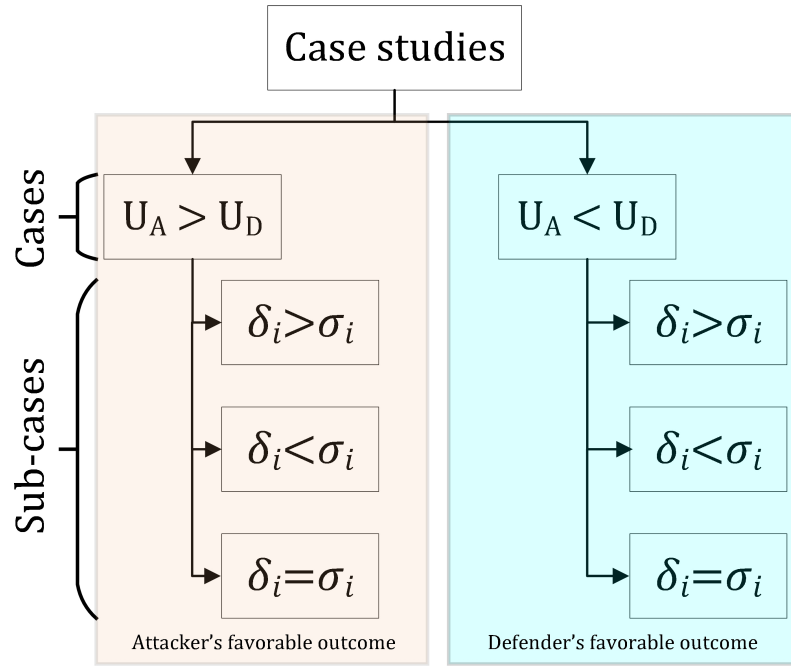


Figure 4.17. Case studies for the solution of the repeated game using reinforcement learning on IEEE 39 bus system.

Figure 4.17 is showing different case studies that are conducted under different conditions to solve the repeated game using reinforcement learning between the adversaries of the electric power grid. Sub-cases are the different strengths of the attacker and the defender. Table 4.11, shows some general settings regarding the game that are commonly used in different case studies. There are 10 independent rounds conducted for the game simulation, and each round consists 5000 runs. The initial exploration probability is set to 0.8. It gradually drops to a very small, and positive final value,  $\epsilon_f$  until the end. A very small and positive value of  $\epsilon_f$  ensures the convergence to the optimal policy and still avoid

local minimum point.

Table 4.11. Value of the parameters

Parameter	Value	Parameter	Value
Total branches	46	$\epsilon$	0.8
$B_A$	5 actions	$\gamma$	0.9
$B_D$	5 actions	$F$	5000
$N(S_A)$	10	Avg	10
$N(S_D)$	10	TLC	6150 MW
$C_a$	$0.01 \times TLC$	$C_m$	$0.01 \times TLC$

Table 4.12 is showing the attacker's and the defender's action sets. These action sets are containing 10 different transmission lines which are selected randomly. Targets are selected from these target sets according to the budget.

Table 4.12. The attacker's and the defender's target sets containing the transmission lines from IEEE 39 branch system. Each of the target sets have 10 transmission lines.

Parameter	Line index
Attacker's set, $S_A$	[18 20 5 13 6 33 3 26 43 37]
Defender's set, $S_D$	[19 21 40 46 36 1 32 10 45 29]

Next we analyze the different outcomes of this game in different conditions for the attacker's and the defender's favorable outcome.

**Case I ( $U_A > U_I$ ):** In this subsection we conduct the game to find the outcome in favor of the attacker. Three different conditions are applied for different attacker's and defender's strengths.

$\delta_i < \sigma_i$ : In this condition, the game is solved in favor of the attacker, and the attacker has higher probability of attacking (higher strength) than the defender. Table 4.13 is showing the optimal actions for the attacker and the defender with their associated probabilities,

and Q-values. The first actions (at time step 1), for the attacker and the defender are attacking transmission line 43 and defending transmission line 36. These actions have the highest probability (0.92) of selection among the other available actions.

Table 4.13. Attack and defense actions with their associated probabilities, and Q-value

Time step	Attack line index	Defense line index	Probability	Q-value
1	43	36	0.92	0.66
2	26	21	0.94	0.73
3	20	40	0.95	0.81
4	5	10	0.97	0.90
5	13	1	0.98	1

The updating of the probabilities of the optimal actions throughout the 5000 runs are shown in the Figure 4.18. The oscillations inside the dotted green box shows the probability update during the random actions selection.

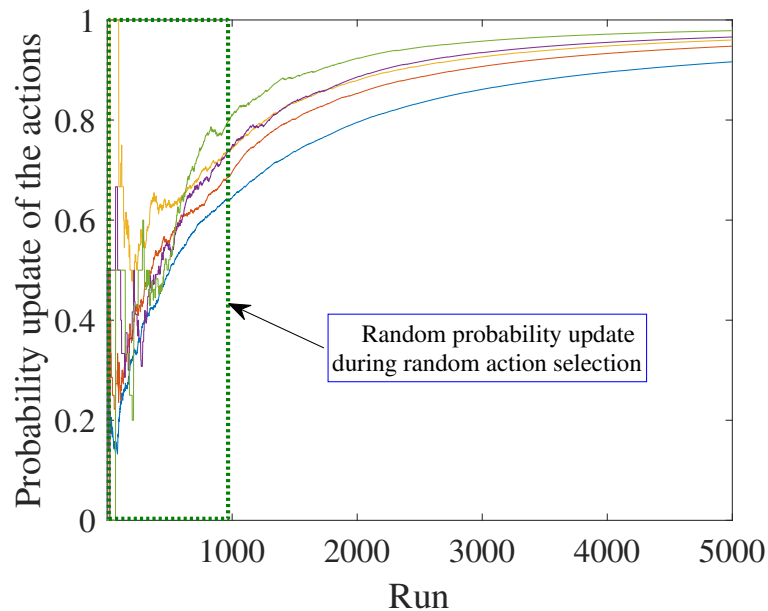


Figure 4.18. Probability update for the optimal actions throughout the 5000 runs.

Figure 4.19, is showing the action selection probabilities of all the possible actions for

selecting the second actions. The right most probability is showing the probability for selecting action 26 and 21 for the attacker and the defender.

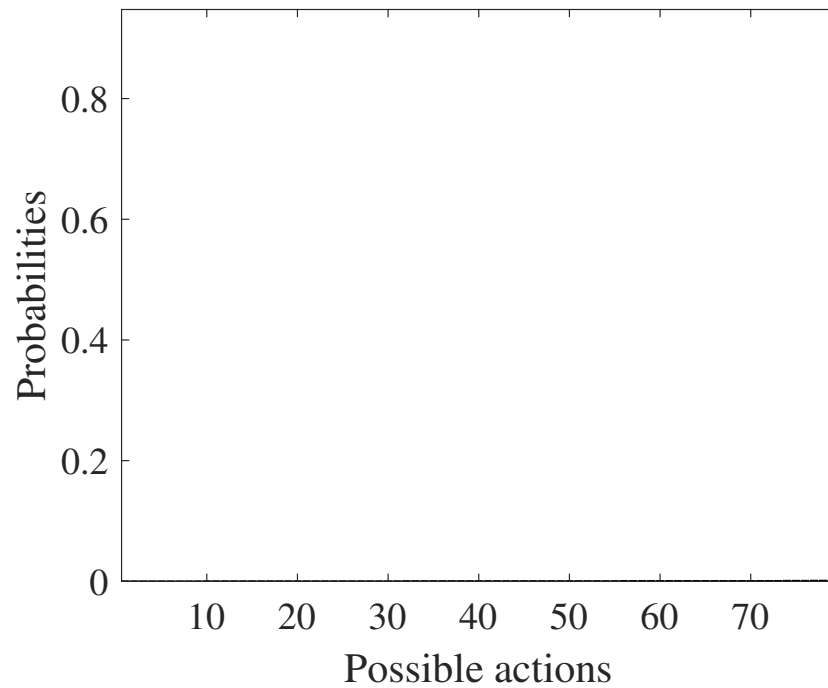


Figure 4.19. Probability of all the possible actions for selecting the second actions.

Table 4.14. Probabilities of all the possible actions for selecting the second actions

Attack line index	Defense line index	Probability
26	21	0.941
20	1	0.001
26	10	0.001
6	46	0.001
3	40	0.001
20	45	0.001
...	...	...
3	10	0.000
20	46	0.002

Table 4.14 shows the probabilities of all the possible actions for second action selection. These probabilities help the players to select their actions. Actions 26 and 21 have

the highest probability of 0.94. In any case, if the actions with highest probability are inaccessible or not available, the actions with second highest probability will be selected by the players. In this case, actions 20 and 46 will be selected as the action by the attacker and the defender, respectively.

$\delta_i > \sigma_i$ : In this condition, the defender has the higher strength than the attacker. Having higher strength of the attacker, there are some combinations and sequence of outcome where the attacker will be favored. Here, we are going to find out the attacker's favorable outcome despite of the defender's higher strength.

Table 4.15. Optimal action sequences for the attacker and the defender with their probability, and Q-values.

Time step	Attack action	Defense action	Probability	Q-value
1	6	1	0.92	0.66
2	5	40	0.92	0.73
3	26	19	0.97	0.81
4	20	36	0.98	0.90
5	37	10	0.98	1

Table 4.15 is showing the optimal action sequences in favor of the attacker while the defender has the higher strength than the attacker. Table 4.16 is showing the probabilities of all the possible actions while selecting the second action 5 and 40 for the attacker and the defender.

Table 4.16. Probabilities of all the possible actions for selecting the second actions

Attack action	Defense action	Probability
18	45	0.0323
43	19	0.0011
20	32	0.0015
5	21	0.0002
43	46	0.0009
18	40	0.0020
...	...	...
5	40	0.9153
33	32	0.0011

From Table 4.16, we can see that, Attack action 5 and defense action 40 has the highest probability among the available actions for selecting the second actions. So these actions are selected as the second action after attacking transmission line 6 and defending transmission line 1. But in any circumstances, if 5 and 40 is not available or inaccessible, the attacker and the defender are going to take the actions with the next highest probability. The actions with the next highest probability are 18 and 45. This way, if inaccessible, actions with next highest probability will be selected.

$\delta_i = \sigma_i$ : This conditions provide the attacker and the defender with equal strength. Which means, the attacker and the defender has the equal probability of attacking and defending (0.5). Having, equal strength, the attacker has some optimal actions for which the outcome will be favorable for the attacker.

Table 4.17. Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the attacker, and the attacker and the defender has equal strength (probability).

Time step	Attack action	Defense action	Probability	Q-value
1	33	40	0.9178	0.66
2	18	21	0.9481	0.73
3	26	46	0.9618	0.81
4	20	1	0.9680	0.90
5	43	19	0.9780	1

Table 4.17, provides the attack and defense actions with their probabilities and Q-values. The Q-values in this table shows that how the rewards are discounted for finding the optimal actions.

Table 4.18. Probabilities of all the possible actions for selecting the second actions in favor of the attacker. The attacker and the defender both have the equal probability or strength.

Attack action	Defense action	Probability
18	21	0.94810
5	1	0.00044
6	1	0.00110
37	19	0.00065
43	19	0.00240
13	21	0.00110
...	...	...
13	45	0.00022
3	10	0.00022

Table 4.18, provides the possible actions for the attacker and the defender while choosing the second action After attacking transmission line 33 and defending 40, according to the highest probability, transmission line 18 and 21 will be selected for the purpose of attacking and defending, respectively. In case of inaccessibility of these transmission lines, transmission lines 43 and 19 will be selected, as they have the second highest probability.

**Case II ( $U_A < U_D$ ):** Similar to Case I, we conduct the game with different attacker's and defender's strengths. But in this case, we define the conditions in a way that all the outcomes or policies goes in favor of the defender. After conducting the game for aforementioned scenarios, we found different attack and defense action policies with their associated probabilities.

$\delta_i < \sigma_i$ : For the first condition in this case study, we will get the optimal actions in favor of the defender. The attacker has higher strength than the defender.

Table 4.19. Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the defender, and the attacker have higher strength (probability of 0.8) then the defender (probability of 0.2).

Time step	Attack action	Defense action	Probability	Q-value
1	3	40	0.73	0.66
2	43	21	0.99	0.73
3	33	32	0.99	0.81
4	6	19	0.99	0.90
5	5	29	0.99	1

Table 4.19, provides the optimal action sequence in favor of the defender ( $U_D > U_A$ ). The probabilities in the fourth column provides the probability of associated actions selection. And the last column provides the Q-value that are discounted rewards in the process of learning the optimal actions.

Table 4.20. Probabilities of all the possible actions for selecting the second actions in favor of the defender. The attacker has higher strength (probability of 0.8) than the defender (probability of 0.2).

Attack action	Defense action	Probability
43	21	0.99128
13	10	0.00055
13	45	0.00027
5	36	0.00027
5	29	0.00055
5	45	0.00055
...	...	...
5	32	0.00055
37	46	0.00027

Table 4.20, gives the probability of all the available actions for selecting the second actions ( At time step 2, attack action on transmission line 43 and the defense action on transmission line 21 have the highest probability (0.99). If these lines are not available, lines with next highest probabilities are selected. If there are multiple actions with same probability, then the actions are taken randomly among them.

$\delta_i > \sigma_i$ : In this subsection we find the optimal action sequences in favor of the defender. Here, the defender has higher strength than the attacker. Table 4.21, shows the optimal actions in favor of the defender where the defender has the higher strength (probability of 0.8) then the attacker (probability of 0.2). The first column is providing the time steps of the attack and the defense actions. The second and third column is providing the actions for the attacker and the defender (associated with that time step). The fourth column is providing the probabilities of these actions to be selected among all the possible actions. The fifth column is showing the Q-values associated with those actions.

Table 4.21. Optimal action sequences for the players with their probabilities, and Q-values. These outcomes are in favor of the defender, and the attacker have higher strength (probability of 0.8) then the defender (probability of 0.2).

Time step	Attack line index	Defense line index	Probability	Q-value
1	5	40	0.92	0.66
2	43	46	0.95	0.73
3	37	10	0.94	0.81
4	33	29	0.97	0.90
5	20	21	0.98	1

Table 4.22. Probabilities of all the possible actions for selecting the second actions in favor of the defender. The defender has higher strength (probability of 0.8) than the attacker (probability of 0.2).

Attack action	Defense action	Probability
43	46	0.9521
26	36	0.0009
3	29	0.0009
26	21	0.0009
43	29	0.0007
33	29	0.0009
...	...	...
18	32	0.0011
3	21	0.0013

Table 4.22, is providing all the possible actions for selecting the second actions when  $U_D > U_A$  and  $\delta_i > \sigma_i$ . Clearly, from the Table 4.22, Attack action 43 and defense action 46 have the highest probability to be selected. Like the other cases and conditions, if they are not accessible, transmission line 3 and 21 will be selected as the attack and the defense actions, respectively, as they have the second highest probability.

$\delta_i = \sigma_i$ : Now in this condition, both the attacker and the defender has equal strength.

The outcome of the game will be in favor of the defender.

Table 4.23. Optimal action sequences for the attacker and the defender with their probability, and Q-values. These outcomes are in favor of the defender, and the attacker and the defender have equal strength (probability of 0.5).

Time step	Attack action	Defense action	Probability	Q-value
1	33	40	0.9178	0.66
2	18	21	0.9481	0.73
3	26	46	0.9618	0.81
4	20	1	0.9680	0.90
5	43	19	0.9780	1

Table 4.23, is providing the optimal actions in favor of the defender, when both of the players have equal strength. This table has also the same format of providing the results like we did in Table 4.21, 4.19, etc.

Table 4.24. Probabilities of all the possible actions for selecting the second actions in favor of the defender. Both the players have equal strength (probability of 0.5).

Attack action	Defense action	Probability
3	36	0.94520
13	19	0.00220
5	1	0.00150
18	21	0.00130
3	46	0.00065
37	46	0.00087
...	...	...
20	46	0.00022
33	10	0.00022

After selecting transmission line 33 and 40 as the first attack and defense actions, similar to the other cases and conditions, actions with highest probability (transmission line 3 and 36 has the highest probability of 0.94520) will be selected. In case of unavailability or inaccessibility actions will be selected according to their probability order.

#### 4.3.0.4 Analysis of impact on the power system

Apart from the game theoretical analysis, in this section we analyze the impacts of these attacks in the power system (e.g., IEEE 39 bus system). The impacts are analyzed for the first case and first condition, where  $U_A > U_I$  and  $\delta_i < \sigma_i$ . The attack in the power system can be illustrated in a time-scale in Figure 4.20.

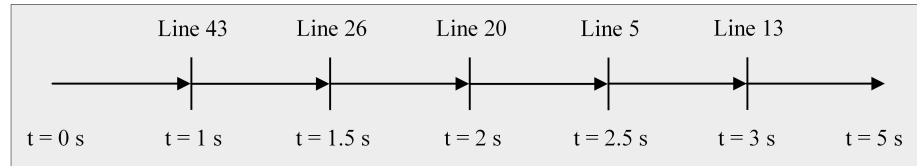


Figure 4.20. The simulation is conducted for five seconds. Starting from  $t = 0$ , the first attack on line 43 is triggered at 1 seconds. Next, line 26, 20, 5, and 13 is triggered at 1.5, 2, 2.5, and 3 seconds.

For simulation purpose, we assume that the attacks are conducted with a time gap of 0.5 seconds. The simulation starts with normal operating condition, where no transmission line is attacked at  $t = 0$  second. After initiating all the transmission line switching attacks, the simulation ends at  $t = 5$  seconds. To assess the disturbances in the system caused by the attacks, we consider the voltage violation of the system elements. There are some defined limits for the voltage violation used in the existing research works [140]–[143], [166]. In our case, we consider the limit of voltage violation is from 0.9 to 1.1 voltage per unit. The per unit voltages at the violated generator 2 and 3 are shown in Figure 4.21(a).

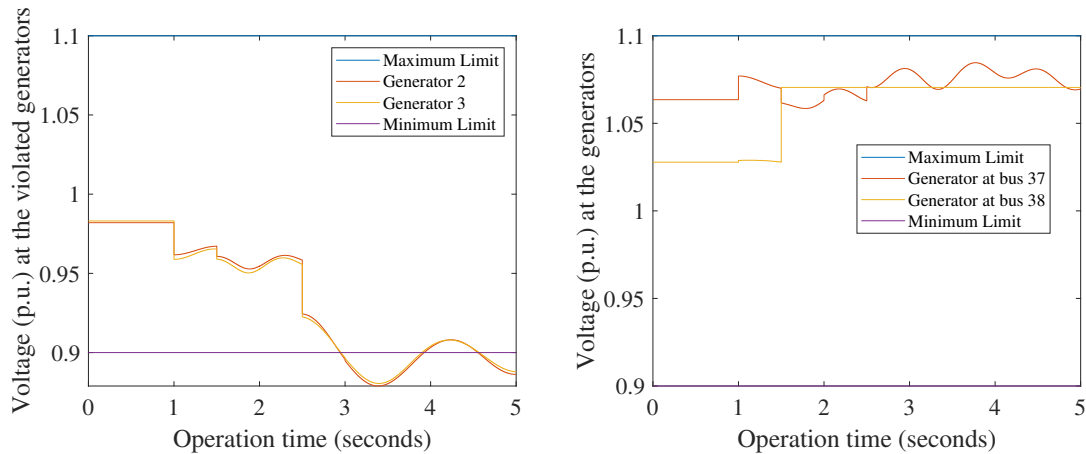


Figure 4.21. Per unit voltages (magnitudes) at the (a) violated generators (generators 2 and 3) during the attack, and (b) vulnerable generators during the attack.

From Figure 4.21(a), we can see that these generator voltages drop below the lower limit (0.9). So these generators violate the voltage (p.u.) limits. Figure 4.21(b), shows the generator voltages (p.u.) of generator at bus 37 and bus 38. These bus voltages are close to the upper limit of the generator voltages (1.1 p.u.). So these generators are vulnerable to violation, because any minor changes in the load or other minor disturbances could shoot these generator voltages above the higher limit.

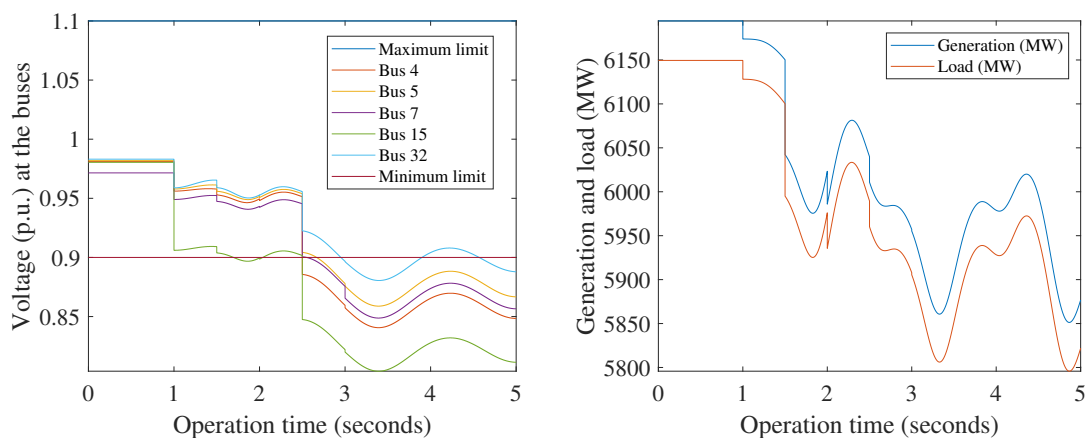


Figure 4.22. (a) Some violated bus voltages (p.u.) and (b) generation and load loss for the whole system during the attack.

Figure 4.22(a), is showing the voltages (p.u.) of bus 4, 5, 7, 15, and 32. We can see

that, these voltages drop below the minimum limit after the attack at 2.5 seconds. Figure 4.22(b), shows the generation loss and load loss during the attack.

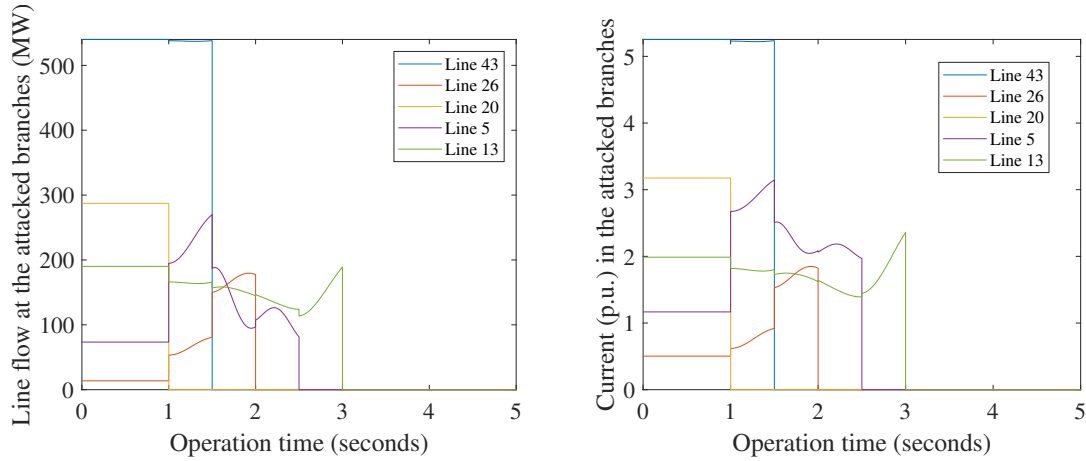


Figure 4.23. (a) Power flow(MW) at the target transmission lines and (b) current flow at the target transmission lines.

Figure 4.23(a), shows the power flow (MW) through the target transmission lines. The line flows drop to zero when they are attacked. Figure 4.23(b), shows the current flow (Amp) through the target transmission lines. The current flow drops to zero when they are attacked.

#### 4.4 Summary

In this chapter, we propose a game-theory based attack-defense repeated game model to analyze the behavior of the attacker and the defender for different probabilities in the smart power grid. Simulation results and the observations give the optimal strategies that the defender should apply in the real-life defensive scheme of power system security. The mixed strategy attack-defense probabilities also provide clear insight about the strategies, while the game theory model gives better performances than the AM model. The budget constraints provide realistic limitations to the attacker's and the defender's action selection.

Moreover, the error and missing detection indicate the performance of the attack detection system in the power grid.

Later in this chapter we propose a novel and effective solution for this game. The solution is proposed based on reinforcement learning algorithm. The learned attacker's and defender's optimal strategies give significant information about the critical transmission lines of a CPPS. Case *I* finds the optimal action sequences for both the players in favor of the attacker for different players' strength. Case *II* provides the optimal action sequences for both the players in favor of the defender for different strengths. These case studies also provide alternative action choices with their probabilities for other possible actions. Section 4.3.0.4 shows the illustration of the attack impacts on the simulated physical system. Generators, and buses that suffer from voltage violation from the attack, are identified. From these case studies, we can suggest the optimal action choices to resolve the game in the attacker's favor or in the defender's favor regardless of their strengths. So, these action sets will help the authorities defend the transmission lines more effectively and efficiently. Moreover, the defense actions are executed against the individual attack actions, which reduces resource consumption of the defensive action schemes.

## CHAPTER 5 Deep learning for adversary game

## NOMENCLATURE

$Q$	Quality of state $s$ , associated with action $a$ and $d$ .
$s$	State of the power system environment.
$a$	Action associated with state $s$ taken by the attacker.
$\alpha$	Learning rate of the Q-learning framework.
$r$	Reward associated with state $s$ and action $a$ and $d$ .
$\gamma$	Discount factor.
$D$	Replay memory of the deep learning framework.
$N$	Size of the replay memory in the deep learning framework.
$R$	Reward assigned for action $a$ and $d$ .
$t$	Time step for action execution.
$\theta$	Weights of the neural network.
$n$	Size of the input to the deep network.
$l$	Line index of agent 1's target set.
$m$	Line index of agent 2's target set.
$S_1$	Agent 1's target set.
$S_2$	Agent 2's target set.
$X$	Elements from the agent's target sets such that $X = \{l, m, \dots\}$ .
$g$	Activation function of the deep Q-network.
$\Delta w$	Changes in the weight
$\eta$	Learning rate of the neural network.
$E$	Squared error in the deep Q-network.
$\epsilon$	Exploration rate.
$Y'$	The next state of parameter $Y$ .
$\mathbb{R}$	Set of real numbers.
$x$	Input of the activation function for deep Q-network.

## 5.1 Chapter overview

Data dimensionality has always been challenging to solve different types of modern infrastructural problems. In real life, the power systems are comprised of several hundreds of thousands buses and branches (transmission lines). Identifying vulnerable elements from these large scale power systems is very time consuming and computationally expensive. Deep learning techniques can address this issue and help the power system planner to identify vulnerable components from large scale power systems. Deep learning technique is also capable of providing faster convergence than the traditional machine learning techniques.

The necessity of using learning techniques in the adversarial game in power system is already discussed in the previous chapters. Deep learning techniques can increase the efficiency and speed of the learning. In our previous work [105], we implemented a multistage sequential game using a reinforcement learning algorithm (Q-learning) to find the optimal action strategies given certain attack objectives. Deep reinforcement learning is being used in different complex game theory environments to reduce the computational complexities. In [167], the authors introduced a deep cognitive hierarchies incorporating joint-policy correlation for multi agent systems. In [168], the authors modeled basic deep learning tasks as strategic games. They proposed Bregman based algorithm for interactive deep generative adversarial networks. The authors in [169], proposed a secure mobile crowd-sensing game utilizing deep reinforcement learning techniques. In [170], the authors adopted deep reinforcement learning for green security game with online information. The authors in [171] used deep reinforcement learning techniques to play chase game. In [172], the authors

discussed about the management game architecture for deep reinforcement learning. Apart from these, deep learning techniques has been used to solve different game theory problems such as atatri, alpha go, etc. [173]–[180]. Except these game theory problems, deep learning based intelligent mechanism has been developed for real time detection of false data injection attacks in smart grid [181]. In [182], the authors used deep learning based game theoretical energy management for energy internet with big data-based renewable power forecasting. The authors in [183], used deep reinforcement learning technique for security and safety in autonomous vehicle systems. In [184], the authors used deep transfer Q-learning with virtual leader-follower for supply demand stackleberg game of smart grid. In the existing literature, deep learning based game theoretic problem has been rarely solved for power system adversarial network. In terms of vulnerability identification, deep learning technique can be used along with the game theory to find the critical components of a power system, optimal strategies for multiagent systems.

Motivated by the limitations of the aforementioned literature, we have the following contributions:

- To minimize the security threats and reduce the damages from the security threats, we design a adversarial multistage gaming framework with a co-simulation platform leveraging the state-of-the-art machine learning technique. The co-simulation platform utilizes the interoperability feature of the grid to accomplish learning and gaming tasks and provides with optimal action strategies.
- The newly designed adversarial gaming framework is employed in a multiagent systems and capable of solving the curse of dimensionality problem for large scale crit-

ical infrastructures like power systems. We conducted a comparative study between the deep learning based adversarial gaming and Q-learning based multistage gaming. The proposed deep learning based gaming approach improves the speed of learning/-training convergence. We compared the speed of convergence of learning of deep learning based adversary game and Q-learning based multistage game and showed that, agent in deep learning based adversary game converges faster under similar environment settings.

The rest of this chapter is organized as follows. Subsection 5.2 provides the explanation of the overall block diagram and flowchart, brief discussion about the developed co-simulation framework, theoretical introduction of Deep-Q-learning and design parameters. Subsection 5.4 provides the simulation setup and different case studies along with the discussion. Subsection 5.5 concludes the chapter with summarizing the outcomes of this chapter and the future opportunities along this direction.

## 5.2 Proposed Framework

In this section, we present the overall block diagram and the designed co-simulation framework in detail.

### 5.2.1 Overall block diagram

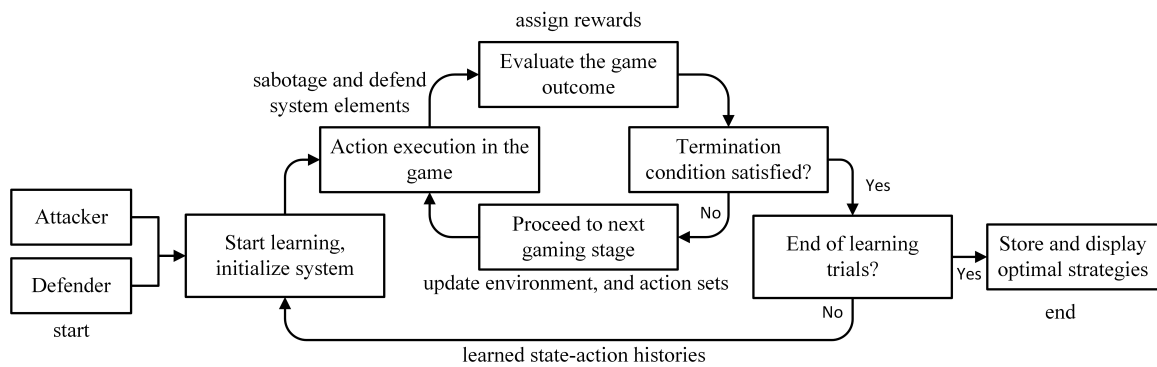


Figure 5.1. Attacker - defender two player learning and gaming interaction in power system adversarial network.

Figure 5.1 represents the overall block diagram of the gaming process involving learning technique and gaming between the agents in power system. Here the agents are the attackers and the defenders. During the learning, previously learned state action histories are transferred to the next learning stages. During the gaming process, the attacker and the defender keep executing actions till their objectives are reached or their allocated budget is totally consumed. At the end, optimal strategies for both players are stored and displayed. We already know from [185], in reinforcement learning, the power system is considered as the environment and the attacker and the defender acts as the agents. Based on the agents' actions in the environment, reward is assigned. So, the main gaming and learning process start by initializing the system to it's steady state by conducting pre-contingency power flow. The adversaries will select their actions from their associated target sets and execute those actions in the power system (environment). Here executing the action represents disconnecting the transmission lines from the system. After executing the actions based on the threat and attack model [95], the learning system will evaluate the actions and

assign rewards. Then the agents will check if the termination condition is satisfied. The termination condition may include the total consumption of the available resource of the agents, or fulfillment of the agents' objectives. If the termination condition is not satisfied, the agents will proceed to the next gaming stage. In the next gaming stage, the agents will again execute actions and evaluate the executed actions' outcome. If satisfied, the agent will check if the learning is complete or not. If the learning is not complete, the agents will reinitialize the system and continue gaming again. If the learning is complete, the agents will quit learning and gaming and, store and display the learned strategies. This process will be repeated multiple times to achieve multiple optimal strategies.

### 5.2.2 Co-simulation framework

In order to implement this game and learning, we introduce a co-simulation framework incorporating MATLAB and PYTHON. The learning framework is built on PYTHON and the power system environment is working on MATLAB.

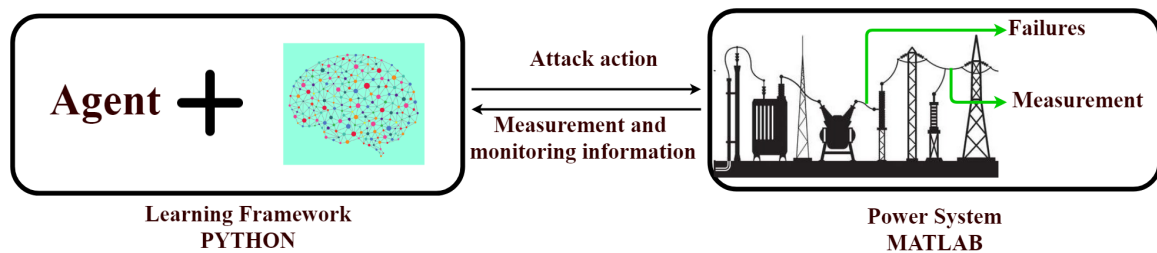


Figure 5.2. Co-simulation framework for learning and gaming in adversarial game in power system. The agent is learning with the help of deep learning technique executing actions in the power systems.

Figure 5.2, represents an abstract visualization of the co-simulation framework for learning and gaming. The learning framework sends the action index/indices to the power system in MATLAB and the power system evaluates the execution of that action and sends

necessary information to the learning framework in PYTHON. Thus, these two platforms keep communicating each other until gaming and learning is completed. We use Deep Q-network to develop this learning framework. Deep Q-learning is a subset of deep reinforcement learning. In [185], we used Q-learning algorithm to implement the gaming environment and provide solution as the optimal strategies to the players. In Deep Q-learning, instead of building a Q-table (tabular method) a neural network is implemented which takes a state and approximates the Q-values for each action based on that state.

### 5.3 Implementation and Parameter Designs

#### 5.3.1 Deep Q-learning

A deep Q-network takes a stack of elements (targets) as input. These targets pass through the network, and output a vector of Q-values for each action possible in the given state. We have to identify the largest Q-value of this vector which will lead us to our best action possible. In the beginning, the agent takes random actions to explore and performs very poor reaching the objective. Over time, it starts to map states with best actions to do.

Reinforcement learning agent evaluates its learning performance by the outcomes it generates. The outcome is goal oriented, and its target is to learn sequences of actions that will help the agent to acquire its objective or maximize its objective function. Q-value maps any state-action pair to their reward.

Experience replay [186] was developed in order to speed up the convergence of RL as well as to train the learning agents in more complex environments.

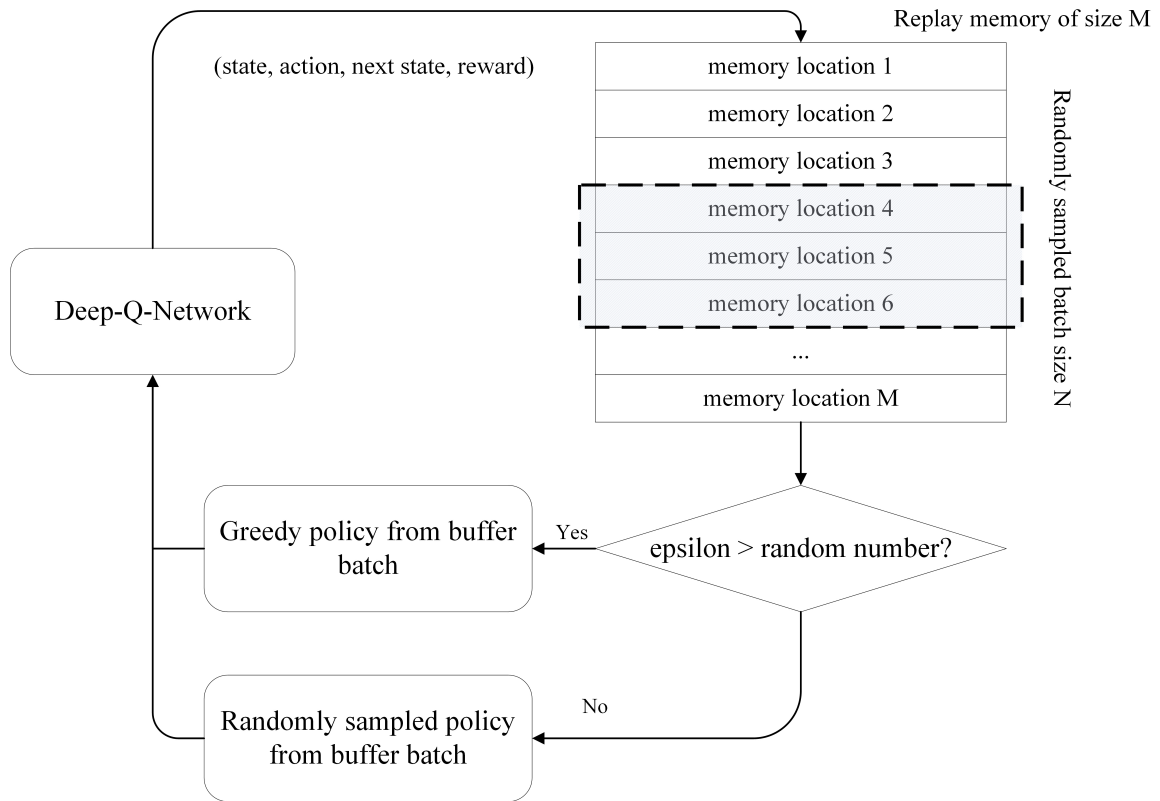


Figure 5.3. Experience replay working mechanism in deep-Q-network. DQN agent stores tuple of state, action, next state and reward in the replay buffer. From the Buffer the agent takes enough sample which are then updated based on randomly sampling from the batch or following greedy policy.

Figure 5.3 represents the working mechanism of an experience replay in a deep-Q-network. The input might be a sequence of experiences which can be highly correlated. So, if the neural network is trained in sequential order, there is a risk for the agent to be influenced by the effect of this correlation. This problem can be solved by sampling random experiences from replay buffer. So we design the deep Q-network with deep neural network and experience replay in order to break the correlations between the input samples. This reduces the probability of the action values from diverging or oscillating severely. An experience is defined as a tuple of  $(s, a, s', r)$  which means that the current state  $s$  if an action  $a$  is executed then the agent moves towards state  $s'$  and gains the reward  $r$  from the

environment. In experience replay, the learning agent takes the past experiences from the replay memory. As a result, the convergence gets faster due to knowledge gained from the good or the bad experiences. During the training process, a batch of training samples is taken from the replay memory and the experiences from that particular batch are used to train the agent. The batch might have useful or random experiences which is chosen based on the learning strategy of the agent. However, the agent should follow the policy that was used to create the experiences. Any change in the learning or training policy of the agent will make the choice of experiences go harmful in successive trials. ER can dramatically reduce the time to learn but increases the computation and memory to store experiences which are generally cheaper than the agent's interaction with the environment (bootstrapping). ER is not a learning algorithm but is combined with other techniques to form the system. In this research, experience replay is used with deep Q-learning.

While implementing the deep-neural-network, the use of experience replay helps to avoid forgetting previous experiences and reduce the correlations between experiences [187]. One of the critical issues is the variability of the weights, as there is high correlations between the actions and states. The main problem is that we provide sequential samples from interactions with the power system environment to our neural network. Every time step, it forgets the previous experiences as it is replaced or overwritten by the new experiences. In this case, making use of the previous experiences multiple times will be more efficient for the learning agent. To solve this issue we use a replay buffer which stores experience tuples during the interactions with the environment. Then, we sample a batch of experience tuples to feed to our neural network for updating. Another problem regarding sampling experiences is that, every action affects the next state.

---

**Algorithm 4: Deep-Q-network for gaming and learning**


---

**Input** : Test case information, Playing agents' (attackers and defenders) target sets,  
allocated budgets, costs of actions

- 1 Initialize co-simulation platform for gaming and learning;
- 2 Initialize replay memory  $D$  to capacity  $N$ ;
- 3 Initialize the network with random weights  $\theta$ ;
- 4 **for** *each episode* **do**
  - 5     initialize starting state for attacking and defending agents;
  - 6     initialize target action-value function  $Q^*$  with weights  $\theta^- = \theta$ ;
  - 7     **for** *each time step* **do**
    - 8         select and execute action via exploration or exploitation;
    - 9         observe reward and next state and store experiences in replay memory;
    - 10        sample random batch of transitions  $(s_t, a_t, R_t, s_{t+1})$  from replay memory  $D$ ;
    - 11        pass the batch of states to policy network;
    - 12        **if** *episode terminates at step  $t + 1$*  **then**
      - 13            set  $R_t$ ;
    - 14        **else**
      - 15            set  $R_t + \gamma \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1}; \theta^-)$ ;
    - 16        **end**
  - 17     **end**
  - 18     calculate loss between output Q-values and target Q-value;
  - 19     gradient descent updates weights in the policy network to minimize the loss;
- 20 **end**

**Output:** Number of actions to objective, optimal action strategies

---

Algorithm 4 provides the detailed algorithm of the adversary game adopting deep-Q learning and experience replay. To learn first, we must take some random actions and save the experiences to the replay buffer. Later we will learn from those saved experiences. As a result, the agent will be able to avoid the problem of being fixated on one region of the state space. It will prevent the agent from selecting same actions again and again.

RMSProp is used as the optimizer in this research. In NN, an optimizer tries to minimize the error between the target and obtained output. RMSProp scales all of the parameters inversely proportional to the exponentially decaying square root of the sum of the previous squared gradients [188]. RMSProp is widely used optimizer in NN applications and has been proven as one of the better algorithms for DNN [189]. The Q-value is updated for a given state and action using the following Bellman equation:

$$Q(s, a) = Q(s, a) + \alpha[R(s, a) + \gamma \max_{a'} Q'(s', a') - Q(s, a)] \quad (5.1)$$

The neural network's weights are updated to reduce the error. The error (or loss) is calculated by taking the difference between the  $Q_{target}$  (maximum possible value from the next state) and  $Q_{value}$  (our current prediction of the Q-value). So the change in weights can be derived as:

$$\Delta w = \alpha[(R + \gamma \max_a \hat{Q}(s', a, w)) - \hat{Q}(s, a, w)] \nabla_w \hat{Q}(s, a, w) \quad (5.2)$$

The reward is assigned based on the actions taken by the agents. If the attack is successful, both attacking agents receive positive rewards and if the attack is not successful, both attacking agents receive negative rewards.

### 5.3.2 Design parameters

#### 5.3.2.1 Hyper-parameters

In the proposed gaming problem, we use a deep neural network to approximate the Q-values for a given state. As the input of the deep network, we use set of transmission lines. These set of transmission lines are the target set of the agents. For multi agent systems, we use separate neural networks. If we define the size of the input as  $n$ , then agent 1's target set as

$$S_1 = \{l_1, l_2, \dots, l_n\} \quad (5.3)$$

where  $S_1$  is agent 1's target set, and  $l_n$  represents the  $n^{th}$  element of agent 1's target set. Similarly, agent 2's target set can be defined as follow

$$S_2 = \{m_1, m_2, \dots, m_n\} \quad (5.4)$$

where  $S_2$  represents agent 2's target set, and  $m_n$  represents the  $n^{th}$  element of agent 2's target set. The elements of the target set contains their status. We use boolean representation of the elements to define active or inactive status. If the element in a target set is active, it is stored as 1. If the status is inactive/disconnected, we use 0 to represent that element's selection status.

$$s_t(X) = \begin{cases} 1, & \text{if line } l \text{ is in-service at time } t \\ 0, & \text{if line } l \text{ is out-of-service at time } t \end{cases} \quad (5.5)$$

where,  $t = \{1, 2, \dots\}$  and  $X = \{l, m, \dots\}$ .

For the deep network we used multiple hidden layers and each layers containing multiple hidden neurons. The number of hidden layers and hidden neurons are determined by trial and error. For selecting the number of hidden neurons we tried to select a value that closely

matches to  $2^n$  formula. Here,  $2^n$  represents the size of the input, where  $n$  represents the number of hidden neurons.

#### 5.3.2.2 Activation function

In [190], the definition of activation function is written as, “An activation function is a function  $g : \mathbb{R} \rightarrow \mathbb{R}$  that is differentiable almost everywhere.” Activation function decides whether to activate or deactivate the neuron based on a certain condition. Most of the real world problems are non-linear in nature. Due to the non-linear activation functions, the neural networks now have been able to differentiate such data that cannot be classified linearly in the data space [191]. However, due to the difficulty in training the deep neural networks (DNN), the non-linear activation function has to be selected carefully in order to make the training process go smooth. In our research, considering the problem of deep reinforcement learning, we have used a non-linear activation function, rectified linear unit (ReLU). Some of the existing non-linear activation functions are sigmoid, hyperbolic tangent, etc. ReLU [192] is the most popular activation function in NN, including DNN. Firstly, it is a non-saturated function and secondly, the derivative of ReLU function results in a constant value which solves the vanishing gradient problem for any NN problems. Hence, due to these reasons, ReLU has been applied in our research as well. The ReLU function is defined as,

$$g(x) = \max(0, x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (5.6)$$

The derivative of (5.6) can be written as,

$$g'(x) = \max(0, x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (5.7)$$

As mentioned before, the output of (5.7) is constant if  $x > 0$  due to which the vanishing gradient problem, as observed in tanh and sigmoid activation functions, is addressed. Other than this, ReLU function decreases the computations of NN and converges much faster, it allows the network to observe sparse representation as the negative inputs result is 0 gradient and most importantly, DNN performs the best when trained with ReLU activation function [193], [194].

#### 5.4 Simulation studies

In this section, we will conduct some simulation studies to validate the solution of the game outcome we obtained through learning technique. We conduct three case studies using the proposed co-simulation framework for gaming and learning in power system vulnerability assessment.

##### 5.4.1 Simulation setup

We conducted the simulation in a co-simulation framework where the learning framework is built on python and the power system framework is built on MATLAB. IEEE 39 bus system and IEEE 300 bus systems are used to conduct the case studies. For power system framework, we conducted simulation using MATLAB *R2018a* on a standard PC with an Intel(R) *i7 – 6700* CPU running at 3.40GHz and 24.0 GB RAM. For learning framework, we conducted simulation using Python 3.6.8.

IEEE 39 bus and IEEE 300 bus system have the following configurations:

Test cases	IEEE 39 bus system	IEEE 300 bus system
Total capacity	6150 MW	23740 MW
Transmission lines	46	516
Number of buses	39	300
Number of generators	10	69

The value of the hyper parameters (for learning) are selected as follows. The discount factor,  $\gamma$  is selected as 0.95. The value of learning rate,  $\alpha$  is selected as 0.001. The initial value of exploration rate,  $\epsilon$  is selected as 1 which ensures very high number of random action selection in the beginning. The value of  $\epsilon$  is gradually reduced close to zero to ensure maximum exploitation at the end of the learning.

#### 5.4.2 Case studies

We conduct three case studies here. The first case study shows that, for IEEE 300 bus system how a single attacking agent can find its optimal strategy to reach attack objective. In the second case study, we use IEEE 39 bus system to represent, how two attacking agents can reach their attack objectives by taking combined actions from their associated target sets. In the third case study, we compare the convergence speed of the agents in previously used Q-learning based multistage dynamic game and our proposed deep learning based adversary game.

##### 5.4.2.1 Case I: one attacker, and one defender

In this case study, we let a single attacking agent to attack against a passive defensive agent. For the defender, we pre-define the defense set as  $\{1, 2, 3\}$  and assume that, the defending transmission lines cannot be attacked.

Table 5.1. Parameter details for gaming and learning for one attacker and one defender

Info	Value	Info	Value
Benchmark	IEEE 300 bus system	Total transmission line	516
Input size	125	Total number of runs	30
Target generation loss	5000	Allowed maximum number of actions	5
Memory size	8000	Batch size	256
Total episode	3000	Number of hidden neuron	12
Number of hidden layer	3	Final epsilon	0.0001

Table 5.1, represents the parameter settings of IEEE 300 bus system for use in learning the optimal strategy in gaming. The number of episodes here is 3000. The value of  $\epsilon$  is adjusted (reduced) such a way that after enough exploration it reduces and stops reducing at 0.001 (final value of epsilon). The input size 125 represents, a vector of transmission lines  $\{1, 2, 3, \dots, 125\}$ . The input will be fed to the network as the status of these transmission lines  $\{1, 1, 1, \dots, 1\}$ .

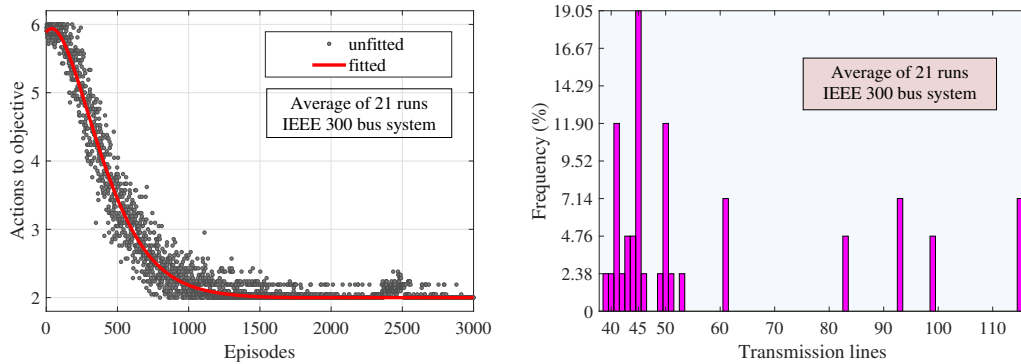


Figure 5.4. (a) Actions to objective (number of attack actions) and (b) frequency of the transmission lines for IEEE 300 bus system.

Figure 5.4(a) represents the convergence of the attacking agent to optimal policy to find minimum number of steps to reach the attack objective. The attack objective here is to cause at least 5000MW generation loss by switching minimum number of transmission lines. The X axis represents the episodes and the Y axis represents the steps to goal (num-

ber of actions required). We used a curve fitting method to show a clear tendency of the agent to convergence. From the figure, we can see that, the agent reaches to optimal policy ( 2 actions) fulfilling the attack objective. It means, the attacking agent will require at least two transmission lines to attack in order to reach his attack objective. We conducted this game 21 times to get different optimal policies. Figure 5.4(b) represents the frequency of the transmission lines that appeared in the optimal policies. Observing the figure, we can see that transmission line 45 is the most frequently used transmission line appeared in the optimal policies. Transmission line 41 and 50 are found as the second most frequently appeared transmission lines in the optimal strategies. This information will help the system planner to enforce additional security measures to the most damaging transmission lines. Additionally, the next highest frequently appeared transmission lines should also be protected in case, the attacker attacks line 41 or 50 instead of attacking transmission line 45.

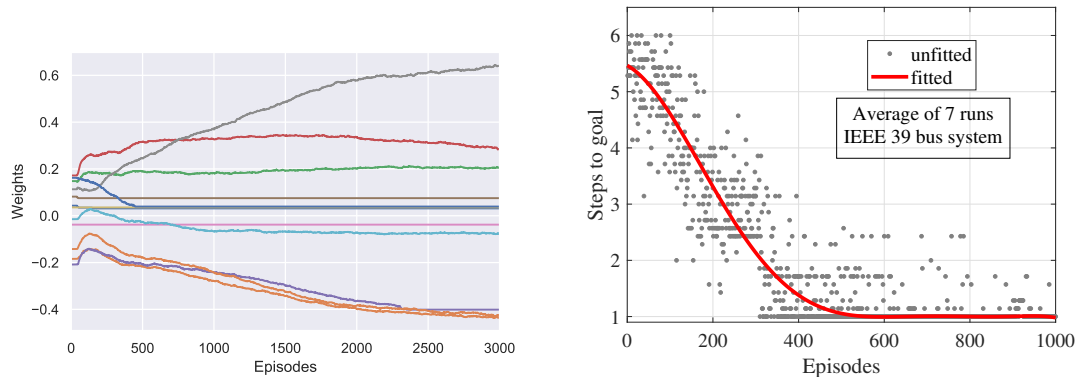


Figure 5.5. (a) Weight update from all input neuron to first hidden layer and (b) convergence of the attacking agents to optimal action strategy (1 transmission line each agent) in IEEE 39 bus system (average of 7 runs).

Figure 5.5(a) represents the weight update from all input neurons to first hidden layer during the learning process.

#### 5.4.2.2 Case II: two attackers, and one defender

In this case study we introduce two attacking agents finding the optimal strategy against one passive defender. For this case study, we use IEEE 39 bus system as the test case.

Table 5.2. Parameter details for gaming and learning for two attacker and one defender

Info	Value	Info	Value
Benchmark	IEEE 39 bus system	Total transmission line	46
Input size	23 (each attacking agent)	Total number of runs	10
Target generation loss	5000 MW (combined)	Allowed maximum number of actions	5
Memory size	6000	Batch size	128
Total episode	1000	Number of hidden neuron	12
Number of hidden layer	2	Final epsilon	0.0001

Table 5.2 presents the information regarding the gaming and learning. Here, we have two attacking agents with target size 23 for each of them. So attacking agent 1's target set is  $\{1, 2, 3, \dots, 23\}$  and the attacking agent 2's target set is  $\{24, 25, 26, \dots, 46\}$ . For this case, we are considering the passive defender's defense scheme as  $\{4, 20, 35\}$

Figure 5.5(b), represents the convergence curve of steps to goal for both the attacking agents to reach attack objective jointly. The agents converge to 1 which means, both of the agents need to take at least one action which jointly will reach to the attack objective. In the figure, we can see some oscillation in the fitted curve which represents the presence of a very small probability of random action selection or exploration to avoid local optimal point.

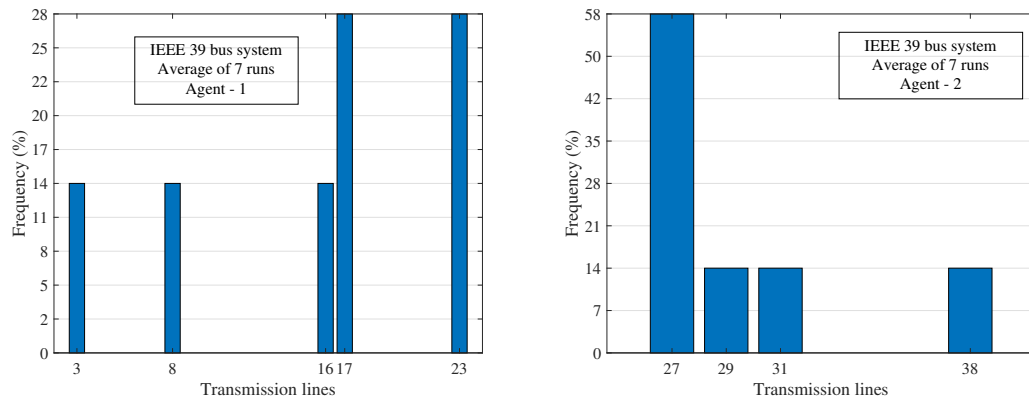


Figure 5.6. (a) and (b) are showing the frequency of the appearance in the optimal target sets for the attacking agent 1 and 2, respectively.

Figure 5.6(a) and 5.6(b) are showing the frequency of the transmission lines that appeared in the optimal strategies for the attacking agent 1 and 2. From the figures, we can see that, transmission line 17 and 23 are the most frequently appeared (hence most vulnerable) transmission line in the target set of attacking agent 1. Similarly, transmission line 27 is the most frequently appeared transmission lines in the attacking agent 2's target set.

#### 5.4.2.3 Case III: performance comparison

Figure 5.7(a) and 5.7(b) represents the steps to goal for attacking agent in Q-learning based multistage sequential game and deep learning based multistage sequential game.

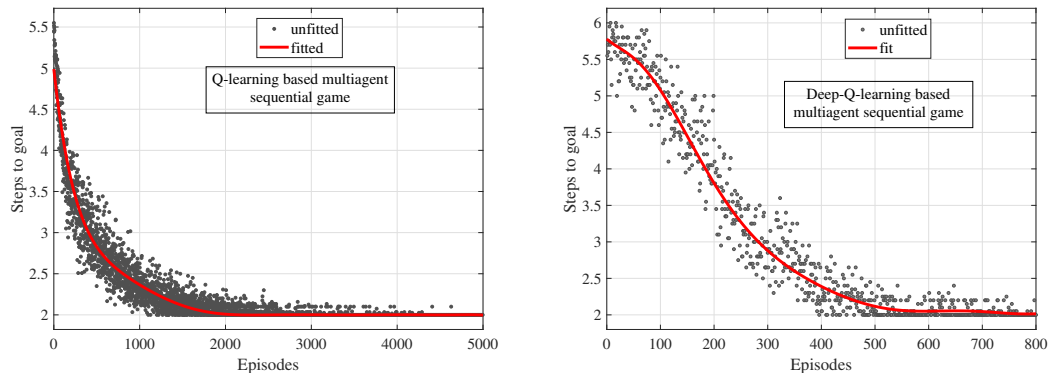


Figure 5.7. Steps to goal for multistage sequential game in IEEE 39 bus system based on (a) Q-learning and (b) deep-Q-learning. In both cases the attacker's target is to cause 5000 MW generation loss.

Both of the agents converge to 2 as the optimal number of steps to goal (causing 5000 MW generation loss). But Q-learning based agent requires 5000 episodes to converge to the optimal number of actions, where deep-learning-based agent requires only 800 episodes to reach to attack objective. Hence this comparison proves that in deep-learning-based multistage sequential game agent converges faster than Q-learning-based multistage sequential game.

## 5.5 Summary

Deep learning techniques are increasing the efficiency of the learning based algorithms and enhancing the problem solving capability in a larger scale. In this chapter we formulated the multistage sequential game theory problem in power system adversary game and proposed the solution using deep-Q-network. In case study 1, we show that, in the presence of a passive defender an attacking agent can learn to select optimal action using deep neural network. In case study 2, we showed for multiagent systems (two attacker, one passive defender), both of the agents successfully learn to find optimal strategy to reach their joint

attack objective. In case study 3, we compare the performance of deep-learning-based adversary game with previously implemented Q-learning based multistage sequential game and showed that, deep learning based agent converges faster to find optimal strategy which successfully fulfill its attack objective. The outcomes of the research of this chapter is submitted and under review for publication.

## CHAPTER 6 Conclusions and opportunities

In this dissertation, we explored some of the critical challenges that power grids are facing. We introduced some new and unconventional methods and techniques to address these challenges. Most of our results have theoretical foundation and applications to practical grid. Some of the framework we developed have applications beyond the power grid network. Exploring these applications and extending some of the research works in this dissertation can be of interest to researchers in power engineering as well as various other technical fields. Some of the future research ideas, extensions and directions are given below:

The threat and attack model adopted in the research works presented in this dissertation is using DC power flow. Study of transient behavior during the threat and attack modeling can reveal some hidden failure and vulnerability of the power system. So far we designed the research experiments using the one-shot, simultaneous and sequential attack scheme. During the sequential attack scheme we considered the order of the attack. But the time to initiate/trigger the subsequent attack has not been identified. Identifying the time sequence of the subsequent attacks could help the power system operators to take more accurate and time sensitive steps.

Traditional and deep learning based solutions have been introduced to the adversarial game theoretic problems for power system. Utilizing partial observable nature of the power system environment could explore some critical vulnerability of power system.

The threat and attack model is implemented in the research works considering that, the attacker has access (by cyber intrusion) to the power systems' control center. Coordinating

a actual cyber intrusion (DDoS, brute-force, phishing, etc.) and physical attack will be state-of-the-art cyber-physical attack defense test-bed.

## LITERATURE CITED

- [1] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 99–107, 2010, ISSN: 1949-3053. DOI: 10.1109/TSG.2010.2046347.
- [2] NIST, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0," *National Institute of Standards and Technology, Special Publication 1108r3, USA 2014*, DOI: <http://dx.doi.org/10.6028/NIST.SP.1108r3>.
- [3] Energy Independence and Security Act. (EISA). 110<sup>th</sup> US Congress,, 2017.
- [4] D. Privitera and L. Li, "Can IoT Devices Be Trusted? An Exploratory Study," in *AMCIS*, 2018.
- [5] C. Tu, X. He, Z. Shuai, and F. Jiang, "Big data issues in smart grid – a review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1099 –1107, 2017, ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2017.05.134>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032117307748>.
- [6] B. Yang, J. Yamazaki, N. Saito, Y. Kokai, and D. Xie, "Big data analytic empowered grid applications — is pmu a big data issue?" In *2015 12th International Conference on the European Energy Market (EEM)*, 2015, pp. 1–4. DOI: 10.1109/EEM.2015.7216718.
- [7] J. A. Harer, L. Y. Kim, R. L. Russell, O. Ozdemir, L. R. Kosta, A. Rangamani, L. H. Hamilton, G. I. Centeno, J. R. Key, P. M. Ellingwood, M. W. McConley, J. M. Oppen, S. P. Chin, and T. Lazovich, "Automated software vulnerability detection with machine learning," *CoRR*, vol. abs/1803.04497, 2018. arXiv: 1803.04497. [Online]. Available: <http://arxiv.org/abs/1803.04497>.
- [8] "2018 Grid Modernization Initiative Peer Review," *Department of Energy*, 2018 (accessed on January 16, 2019). [Online]. Available: <https://www.energy.gov/grid-modernization-initiative-0/2018-grid-modernization-initiative-peer-review>.
- [9] "NSF's 10 Big Ideas," *National Science Foundation*, 2018 (accessed on January 16, 2019). [Online]. Available: [https://www.nsf.gov/news/special\\_reports/big\\_ideas/index.jsp](https://www.nsf.gov/news/special_reports/big_ideas/index.jsp).
- [10] M. Mylrea, S. N. G. Gourisetti, R. Bishop, and M. Johnson, "Keyless signature blockchain infrastructure: Facilitating nerc cip compliance and responding to evolving cyber threats and vulnerabilities to energy infrastructure," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, 2018, pp. 1–9. DOI: 10.1109/TDC.2018.8440380.
- [11] L. Akinmodiro, *Nigeria witnesses first total blackout of 2018*, Available at: <http://www.powerconnect.com/blog/2018/01/nigeria-witnesses-first-total-blackout-2018/>, 2018 (accessed February 2, 2019).

- [12] EATON, *Blackout tracker*, Available at: <http://electricalsector.eaton.com/forms/BlackoutTrackerAnnualReport>, 2017 (accessed November 2, 2018).
- [13] *Hourly downtime tops \$300k for 81% of firms; 33% of enterprises say downtime costs \$1m*, Available at: <https://itic-corp.com/blog/2017/05/hourly-downtime-tops-300k-for-81-of-firms-33-of-enterprises-say-downtime-costs-1m/>, 2017 (accessed April 19, 2019).
- [14] W. Wang and Z. Lu, “Cyber security in the smart grid: Survey and challenges,” *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013, ISSN: 1389-1286. DOI: <https://doi.org/10.1016/j.comnet.2012.12.017>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128613000042>.
- [15] S. Rolley, *A foreign entity has breached the us power grid*, Available at: <http://willcountynews.com/2017/09/09/a-foreign-entity-has-breached-the-us-power-grid/>, 2017 (accessed November 28, 2018).
- [16] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, “Detecting stealthy false data injection using machine learning in smart grid,” *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644–1652, 2017, ISSN: 1932-8184. DOI: 10.1109/JSYST.2014.2341597.
- [17] D. S. Terzi, B. Arslan, and S. Sagiroglu, “Smart grid security evaluation with a big data use case,” in *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*, 2018, pp. 1–6. DOI: 10.1109/CPE.2018.8372536.
- [18] Z. Zheng, Y. Yang, X. Niu, H. Dai, and Y. Zhou, “Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2018, ISSN: 1551-3203. DOI: 10.1109/TII.2017.2785963.
- [19] A. Ferdowsi, A. Sanjab, W. Saad, and N. B. Mandayam, “Game theory for secure critical interdependent gas-power-water infrastructure,” in *2017 Resilience Week (RWS)*, 2017, pp. 184–190. DOI: 10.1109/RWEEK.2017.8088670.
- [20] M. N. Soorki, W. Saad, M. H. Manshaei, and H. Saidi, “Stochastic coalitional games for cooperative random access in m2m communications,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6179–6192, 2017, ISSN: 1536-1276. DOI: 10.1109/TWC.2017.2720658.
- [21] K. Vasudeva, S. Dikmese, I. Güvenç, A. Mehbodniya, W. Saad, and F. Adachi, “Fuzzy logic game-theoretic approach for energy efficient operation in hetnets,” in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 552–557. DOI: 10.1109/ICCW.2017.7962716.

- [22] Y. Zhu, D. Zhao, and X. Li, "Iterative adaptive dynamic programming for solving unknown nonlinear zero-sum game based on online data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 714–725, 2017, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2016.2561300.
- [23] R. Song, F. L. Lewis, and Q. Wei, "Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 704–713, 2017, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2016.2582849.
- [24] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 957–969, 2018, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2016.2638863.
- [25] J. Li, H. Modares, T. Chai, F. L. Lewis, and L. Xie, "Off-policy reinforcement learning for synchronization in multiagent graphical games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2434–2445, 2017, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2016.2609500.
- [26] M. Johnson, R. Kamalapurkar, S. Bhasin, and W. E. Dixon, "Approximate n-player nonzero-sum game solution for an uncertain continuous nonlinear system," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1645–1658, 2015, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2014.2350835.
- [27] Y. Fu and T. Chai, "Online solution of two-player zero-sum games for continuous-time nonlinear systems with completely unknown dynamics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 12, pp. 2577–2587, 2016, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2015.2496299.
- [28] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2018, in press.
- [29] H. Zhang, L. Cui, and Y. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network adp," *IEEE Transactions on Cybernetics*, vol. 43, no. 1, pp. 206–216, 2013.
- [30] R. Yao, S. Huang, K. Sun, F. Liu, X. Zhang, S. Mei, W. Wei, and L. Ding, "Risk assessment of multi-timescale cascading outages based on markovian tree search," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2887–2900, 2017.
- [31] J. Qi, W. Ju, and K. Sun, "Estimating the propagation of interdependent cascading outages with multi-type branching processes," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1212–1223, 2017.

- [32] Y. Zhu, J. Yan, Y. L. Sun, and H. He, "Revealing cascading failure vulnerability in power grids using risk-graph," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3274–3284, 2014.
- [33] S. Poudel, Z. Ni, and W. Sun, "Electrical distance approach for searching vulnerable branches during contingencies," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3373–3382, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2016.2631622.
- [34] Y. Zhu, J. Yan, Y. Tang, Y. L. Sun, and H. He, "Resilience analysis of power grids under the sequential attack," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2340–2354, 2014, ISSN: 1556-6013. DOI: 10.1109/TIFS.2014.2363786.
- [35] J. Yan, H. He, and Y. Sun, "Integrated security analysis on cascading failure in complex networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 451–463, 2014, ISSN: 1556-6013. DOI: 10.1109/TIFS.2014.2299404.
- [36] C. Davis and T. Overbye, "Linear analysis of multiple outage interaction," in *2009 42nd Hawaii International Conference on System Sciences*, 2009, pp. 1–8. DOI: 10.1109/HICSS.2009.291.
- [37] C. M. Davis and T. J. Overbye, "Multiple element contingency screening," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1294–1301, 2011, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2010.2087366.
- [38] M. K. Enns, J. J. Quada, and B. Sackett, "Fast linear contingency analysis," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 4, pp. 783–791, 1982, ISSN: 0018-9510. DOI: 10.1109/TPAS.1982.317142.
- [39] C. Lai and S. H. Low, "The redistribution of power flow in cascading failures," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013, pp. 1037–1044. DOI: 10.1109/Allerton.2013.6736639.
- [40] K. S. Turitsyn and P. A. Kaplunovich, "Fast algorithm for n-2 contingency problem," in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 2161–2166. DOI: 10.1109/HICSS.2013.233.
- [41] P. A. Kaplunovich and K. S. Turitsyn, "Statistical properties and classification of n-2 contingencies in large scale power grids," in *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 2517–2526. DOI: 10.1109/HICSS.2014.316.

- [42] P. D. H. Hines, I. Dobson, E. Cotilla-Sanchez, and M. Eppstein, ““dual graph” and “random chemistry” methods for cascading failure analysis,” in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 2141–2150. DOI: 10.1109/HICSS.2013.1.
- [43] V. Donde, V. Lopez, B. Lesieutre, A. Pinar, C. Yang, and J. Meza, “Severe multiple contingency screening in electric power systems,” *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 406–417, 2008, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2008.919243.
- [44] M. Chertkov, F. Pan, and M. G. Stepanov, “Predicting failures in power grids: The case of static overloads,” *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 162–172, 2011, ISSN: 1949-3053. DOI: 10.1109/TSG.2010.2090912.
- [45] D. Bienstock and A. Verma, “The  $n$ - $k$  problem in power grids: New models, formulations, and numerical experiments,” *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2352–2380, 2010. DOI: 10.1137/08073562X. eprint: <https://doi.org/10.1137/08073562X>. [Online]. Available: <https://doi.org/10.1137/08073562X>.
- [46] 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, no. 1, pp. 200–210, 2017, ISSN: 1556-6013. DOI: 10.1109/TIFS.2016.2607701.
- [47] C. Luo, J. Yang, Y. Sun, J. Yan, and H. He, “Identify critical branches with cascading failure chain statistics and hypertext-induced topic search algorithm,” in *2017 IEEE Power Energy Society General Meeting*, 2017, pp. 1–5. DOI: 10.1109/PESGM.2017.8274213.
- [48] N. V. Tomin, V. G. Kurbatsky, D. N. Sidorov, and A. V. Zhukov, “Machine learning techniques for power system security assessment\*\*this work was supported by the russian scientific foundation under grant no. 14-19-00054 and the 2015 endeavour scholarship and fellowship program.,” *IFAC-PapersOnLine*, vol. 49, no. 27, pp. 445–450, 2016, IFAC Workshop on Control of Transmission and Distribution Smart Grids CTDSG 2016, ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2016.10.773>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896316324028>.
- [49] J. L. Cremer, I. Konstantelos, S. H. Tindemans, and G. Strbac, “Data-driven power system operation: Exploring the balance between cost and risk,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 791–801, 2019, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2018.2867209.

- [50] R. Eskandarpour and A. Khodaei, "Machine learning based power grid outage prediction in response to extreme events," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3315–3316, 2017, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2631895.
- [51] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1773–1786, 2016, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2015.2404803.
- [52] A. Valdes, R. Macwan, and M. Backes, "Anomaly detection in electrical substation circuits via unsupervised machine learning," in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 2016, pp. 500–505. DOI: 10.1109/IRI.2016.74.
- [53] F. Luo, Z. Dong, G. Chen, Y. Xu, K. Meng, Y. Chen, and K. Wong, "Advanced pattern discovery-based fuzzy classification method for power system dynamic security assessment," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 2, pp. 416–426, 2015, ISSN: 1551-3203. DOI: 10.1109/TII.2015.2399698.
- [54] M. Marzband, "Non-cooperative game theory based energy management systems for energy district in the retail market considering der uncertainties," English, *IET Generation, Transmission Distribution*, vol. 10, 2999–3009(10), 12 2016, ISSN: 1751-8687. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2016.0024>.
- [55] M. Kristiansen, M. Korpås, and H. G. Svendsen, "A generic framework for power system flexibility analysis using cooperative game theory," *Applied Energy*, vol. 212, pp. 223 –232, 2018, ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2017.12.062>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261917317774>.
- [56] N. I. Nwulu and X. Xia, "Multi-objective dynamic economic emission dispatch of electric power generation integrated with game theory based demand response programs," *Energy Conversion and Management*, vol. 89, pp. 963 –974, 2015, ISSN: 0196-8904. DOI: <https://doi.org/10.1016/j.enconman.2014.11.001>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0196890414009510>.
- [57] M. H. Ranjbar, M. Kheradmandi, and A. Pirayesh, "A linear game framework for defending power systems against intelligent physical attacks," *IEEE Transactions on Smart Grid*, pp. 1–1, 2019, ISSN: 1949-3053. DOI: 10.1109/TSG.2019.2908083.

- [58] 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, ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2016.02.045>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0142061516303106>.
- [59] 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.
- [60] 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.
- [61] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, "Smart attacks in smart grid communication networks," *IEEE Communications Magazine*, vol. 50, no. 8, 2012.
- [62] 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.
- [63] F. Li and Y. Du, "From alphago to power system ai: What engineers can learn from solving the most complex board game," *IEEE Power and Energy Magazine*, vol. 16, no. 2, pp. 76–84, 2018, ISSN: 1540-7977. DOI: 10.1109/MPE.2017.2779554.
- [64] L. Wei, A. I. Sarwat, W. Saad, and S. Biswas, "Stochastic games for power grid protection against coordinated cyber-physical attacks," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 684–694, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2016.2561266.
- [65] L. Wei, A. I. Sarwat, and W. Saad, "Risk assessment of coordinated cyber-physical attacks against power grids: A stochastic game approach," in *2016 IEEE Industry Applications Society Annual Meeting*, 2016, pp. 1–7. DOI: 10.1109/IAS.2016.7731849.
- [66] A. Sanjab, W. Saad, and T. Basar, "Graph-theoretic framework for unified analysis of observability and data injection attacks in the smart grid," *CoRR*, vol. abs/1801.08951, 2018. arXiv: 1801.08951. [Online]. Available: <http://arxiv.org/abs/1801.08951>.
- [67] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, 2018, in press, ISSN: 1949-3053. DOI: 10.1109/TSG.2018.2790704.
- [68] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

- [69] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.
- [70] M. Abu-Khalaf, F. L. Lewis, and J. Huang, “Neurodynamic programming and zero-sum games for constrained control systems,” *IEEE Transactions on Neural Networks*, vol. 19, no. 7, pp. 1243–1252, 2008.
- [71] O. P. Veloza and F. Santamaria, “Analysis of major blackouts from 2003 to 2015: Classification of incidents and review of main causes,” *The Electricity Journal*, vol. 29, no. 7, pp. 42–49, 2016, ISSN: 1040-6190. DOI: <https://doi.org/10.1016/j.tej.2016.08.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1040619016301191>.
- [72] “U.s.-canada power system outage task force. final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations,” System:238, 2004.
- [73] “North American Blackout.[Online],” 2004, Available: <http://www.snopes.com/photos/space/blackout.asp>.
- [74] 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, no. 6, pp. 4887–4900, 2016, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2518660.
- [75] M. Papic, K. Bell, Y. Chen, I. Dobson, L. Fonte, E. Haq, P. Hines, D. Kirschen, X. Luo, S. S. Miller, N. Samaan, M. Vaiman, M. Varghese, and P. Zhang, “Survey of tools for risk assessment of cascading outages,” 2011, pp. 1–9. DOI: 10.1109/PES.2011.6039371.
- [76] and, and and, “Risk assessment of cascading outages: Methodologies and challenges,” *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 631–641, 2012, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2011.2177868.
- [77] J. Chen, J. S. Thorp, and I. Dobson, “Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model,” *International Journal of Electrical Power Energy Systems*, vol. 27, no. 4, pp. 318–326, 2005, ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2004.12.003>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0142061505000232>.
- [78] D. P. Chassin and C. Posse, “Evaluating north american electric grid reliability using the barabási–albert network model,” *Physica A: Statistical Mechanics and its Applications*, vol. 355, no. 2, pp. 667–677, 2005, ISSN: 0378-4371. DOI: <https://doi.org/10.1016/j.physa.2005.02.051>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378437105002311>.

- [79] H. Xiao and E. M. Yeh, “Cascading link failure in the power grid: A percolation-based analysis,” in *2011 IEEE International Conference on Communications Workshops (ICC)*, 2011, pp. 1–6. DOI: 10.1109/iccw.2011.5963573.
- [80] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, “Complex dynamics of blackouts in power transmission systems,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 14, no. 3, pp. 643–652, 2004. DOI: 10.1063/1.1781391. eprint: <https://doi.org/10.1063/1.1781391>. [Online]. Available: <https://doi.org/10.1063/1.1781391>.
- [81] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, “Critical points and transitions in an electric power transmission model for cascading failure blackouts,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 12, no. 4, pp. 985–994, 2002. DOI: 10.1063/1.1505810. eprint: <https://doi.org/10.1063/1.1505810>. [Online]. Available: <https://doi.org/10.1063/1.1505810>.
- [82] M. Anghel, K. A. Werley, and A. E. Motter, “Stochastic model for power grid dynamics,” in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, 2007, pp. 113–113. DOI: 10.1109/HICSS.2007.500.
- [83] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss, “Cascading failures in spatially-embedded random networks,” *PLOS ONE*, vol. 9, no. 1, pp. 1–13, Jan. 2014. DOI: 10.1371/journal.pone.0084563. [Online]. Available: <https://doi.org/10.1371/journal.pone.0084563>.
- [84] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, “Sensitivity analysis of the power grid vulnerability to large-scale cascading failures,” *SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 3, pp. 33–37, Jan. 2012, ISSN: 0163-5999. DOI: 10.1145/2425248.2425256. [Online]. Available: <http://doi.acm.org/10.1145/2425248.2425256>.
- [85] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, “Power grid vulnerability to geographically correlated failures — analysis and control implications,” in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, 2014, pp. 2634–2642. DOI: 10.1109/INFOCOM.2014.6848211.
- [86] Y. Koç, M. Warnier, P. V. Mieghem, R. E. Kooij, and F. M. Brazier, “The impact of the topology on cascading failures in a power grid model,” *Physica A: Statistical Mechanics and its Applications*, vol. 402, pp. 169–179, 2014, ISSN: 0378-4371. DOI: <https://doi.org/10.1016/j.physa.2014.01.056>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378437114000776>.

- [87] I. Dobson, J. Kim, and K. R. Wierzbicki, "Testing branching process estimators of cascading failure with data from a simulation of transmission line outages," *Risk Analysis*, vol. 30, no. 4, pp. 650–662, DOI: 10.1111/j.1539-6924.2010.01369.x. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.2010.01369.x>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2010.01369.x>.
- [88] H. Ren and I. Dobson, "Using transmission line outage data to estimate cascading failure propagation in an electric power system," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 9, pp. 927–931, 2008, ISSN: 1549-7747. DOI: 10.1109/TCSII.2008.924365.
- [89] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the us power grid," *Safety Science*, vol. 47, no. 10, pp. 1332 –1336, 2009, ISSN: 0925-7535. DOI: <https://doi.org/10.1016/j.ssci.2009.02.002>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925753509000174>.
- [90] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, "Model validation for the august 10, 1996 wscs system outage," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 967–979, 1999, ISSN: 0885-8950. DOI: 10.1109/59.780909.
- [91] M. Bhavaraju and N. Nour, "Trelss: A computer program for transmission reliability evaluation of large-scale systems. volume 1, a summary: Final report," May 1992.
- [92] D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch, "Criticality in a cascading failure blackout model," *International Journal of Electrical Power Energy Systems*, vol. 28, no. 9, pp. 627 –633, 2006, Selection of Papers from 15th Power Systems Computation Conference, 2005, ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2006.03.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0142061506000810>.
- [93] B. Stott, J. Jardim, and O. Alsac, "Dc power flow revisited," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1290–1300, 2009, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2009.2021235.
- [94] R. Fitzmaurice, A. Keane, and M. O'Malley, "Effect of short-term risk-averse dispatch on a complex system model for power systems," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 460–469, 2011, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2010.2050079.
- [95] 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, no. 3, pp. 1698–1705, 2012, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2012.2183624.

- [96] S. Chakrabarti and E. Kyriakides, "Optimal placement of phasor measurement units for power system observability," *Power Systems, IEEE Transactions on*, vol. 23, pp. 1433–1440, Sep. 2008. DOI: 10.1109/TPWRS.2008.922621.
- [97] P. Hines and S. Blumsack, "A centrality measure for electrical networks," Jan. 2008, p. 185. DOI: 10.1109/HICSS.2008.5.
- [98] M. J. Osborne and A. Rubinstein, *A course in game theory*. MIT press, 1994.
- [99] 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, no. 5, pp. 86–105, 2012.
- [100] X. Lu, X. Wang, D. Rimorov, H. Sheng, and G. Joós, "Synchrophasor-based state estimation for voltage stability monitoring in power systems," in *2018 North American Power Symposium (NAPS)*, 2018, pp. 1–6. DOI: 10.1109/NAPS.2018.8600661.
- [101] V. B. Krishna, C. A. Gunter, and W. H. Sanders, "Evaluating detectors on optimal attack vectors that enable electricity theft and der fraud," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 790–805, 2018, ISSN: 1932-4553. DOI: 10.1109/JSTSP.2018.2833749.
- [102] Z. Ding, Y. Xiang, and L. Wang, "Incorporating unidentifiable cyberattacks into power system reliability assessment," in *2018 IEEE Power Energy Society General Meeting (PESGM)*, 2018, pp. 1–5. DOI: 10.1109/PESGM.2018.8585884.
- [103] C. Ten, K. Yamashita, Z. Yang, A. V. Vasilakos, and A. Ginter, "Impact assessment of hypothesized cyberattacks on interconnected bulk power systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4405–4425, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2017.2656068.
- [104] L. Lu, H. J. Liu, and H. Zhu, "Distributed secondary control for isolated microgrids under malicious attacks," in *2016 North American Power Symposium (NAPS)*, 2016, pp. 1–6. DOI: 10.1109/NAPS.2016.7747929.
- [105] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2019, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2018.2885530.
- [106] F. Pourahmadi, M. Fotuhi-Firuzabad, and P. Dehghanian, "Application of game theory in reliability-centered maintenance of electric power systems," *IEEE Transactions on Industry Applications*, vol. 53, no. 2, pp. 936–946, 2017, ISSN: 0093-9994. DOI: 10.1109/TIA.2016.2639454.

- [107] W. Liao, S. Salinas, M. Li, P. Li, and K. A. Loparo, "Cascading failure attacks in the power system: A stochastic game perspective," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2247–2259, 2017, ISSN: 2327-4662. DOI: 10.1109/JIOT.2017.2761353.
- [108] Q. Wang, W. Tai, Y. Tang, M. Ni, and S. You, "A two-layer game theoretical attack-defense model for a false data injection attack against power systems," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 169–177, 2019, ISSN: 0142-0615. DOI: <https://doi.org/10.1016/j.ijepes.2018.07.007>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0142061518312171>.
- [109] K. Wang, M. Du, S. Maharjan, and Y. Sun, "Strategic honeypot game model for distributed denial of service attacks in the smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2474–2482, 2017, ISSN: 1949-3053. DOI: 10.1109/TSG.2017.2670144.
- [110] A. Farraj, E. Hammad, A. A. 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, ISSN: 1949-3053. DOI: 10.1109/TSG.2015.2440095.
- [111] Y. Zhu, J. Yan, Y. Tang, Y. Sun, and H. He, "The sequential attack against power grid networks," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 616–621. DOI: 10.1109/ICC.2014.6883387.
- [112] A. Ashok and M. Govindarasu, "Cyber-physical risk modeling and mitigation for the smart grid using a game-theoretic approach," in *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2015, pp. 1–5. DOI: 10.1109/ISGT.2015.7131842.
- [113] S. Paul and Z. Ni, "A study of linear programming and reinforcement learning for one-shot game in smart grid security," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8. DOI: 10.1109/IJCNN.2018.8489202.
- [114] 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)*, 2017, pp. 1–5. DOI: 10.1109/ISGT.2017.8086078.
- [115] Z. Ni, S. Paul, X. Zhong, and Q. Wei, "A reinforcement learning approach for sequential decision-making process of attacks in smart grid," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–8. DOI: 10.1109/SSCI.2017.8285291.

- [116] K. Mahapatra and N. R. Chaudhuri, "Malicious corruption-resilient wide-area oscillation monitoring using online robust pca," in *2018 IEEE Power Energy Society General Meeting (PESGM)*, 2018, pp. 1–5. DOI: 10.1109/PESGM.2018.8586404.
- [117] M. Korkali, J. G. Veneman, B. F. Tivnan, J. P. Bagrow, and P. D. Hines, "Reducing cascading failure risk by increasing infrastructure network interdependence," *Scientific Reports*, vol. 7, 2017.
- [118] J. Bialek, E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, C. Dent, I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller, *et al.*, "Benchmarking and validation of cascading failure analysis tools," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4887–4900, 2016.
- [119] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [120] 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.
- [121] C. Ramos and C.-C. Liu, "Ai in power systems and energy markets," *IEEE Intelligent Systems*, vol. 26, no. 2, pp. 5–8, 2011.
- [122] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.
- [123] J. Li, X.-Y. Ma, C.-C. Liu, and K. P. Schneider, "Distribution system restoration with microgrids using spanning tree search," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 3021–3029, 2014.
- [124] P. Y. Chen, S. M. Cheng, and K. C. Chen, "Smart attacks in smart grid communication networks," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 24–29, 2012, ISSN: 0163-6804. DOI: 10.1109/MCOM.2012.6257523.
- [125] 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.
- [126] 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, ISSN: 0378-7796. DOI: <https://doi.org/10.1016/j.epsr.2017.05.020>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378779617302092>.
- [127] D. Vrabie and F. Lewis, "Adaptive dynamic programming algorithm for finding online the equilibrium solution of the two-player zero-sum differential game," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–8. DOI: 10.1109/IJCNN.2010.5596754.

- [128] R. Song, F. L. Lewis, and Q. Wei, "Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 704–713, 2017.
- [129] K. G. Vamvoudakis and F. L. Lewis, "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 13, pp. 1460–1483, 2012.
- [130] H. Zhang, Q. Wei, and D. Liu, "An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games," *Automatica*, vol. 47, no. 1, pp. 207–214, 2011.
- [131] H. W. Corley, "Games with vector payoffs," *Journal of Optimization Theory and Applications*, vol. 47, no. 4, pp. 491–498, 1985, ISSN: 1573-2878. DOI: 10.1007/BF00942194. [Online]. Available: <https://doi.org/10.1007/BF00942194>.
- [132] Z. Zhu, J. Tang, S. Lambbotharan, 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)*, 2011, pp. 1205–1210. DOI: 10.1109/GLOCOMW.2011.6162372.
- [133] C. Y. T. Ma, D. K. Y. Yau, X. Lou, and N. S. V. Rao, "Markov game analysis for attack-defense of power networks under possible misinformation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1676–1686, 2013, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2012.2226480.
- [134] J. Hu and M. P. Wellman, "Multiagent reinforcement learning: Theoretical framework and an algorithm," in *Proceedings of the Fifteenth International Conference on Machine Learning*, ser. ICML '98, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 242–250, ISBN: 1-55860-556-8. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645527.657296>.
- [135] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, Aug 2012.
- [136] A. Mitra and W. D. Sudderth, *Discrete Gambling and Stochastic Games*. New York: Springer-Verlag, 1996.
- [137] K. Chatterjee and A. Tarlecki, *Computer Science Logic*. Springer Berlin Heidelberg Press, 2004.
- [138] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," in *2010 43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10. DOI: 10.1109/HICSS.2010.35.

- [139] A. Ashok, A. Hahn, and M. Govindarasu, "Cyber-physical security of wide-area monitoring, protection and control in a smart grid environment," *Journal of Advanced Research*, vol. 5, no. 4, pp. 481–489, 2014, ISSN: 2090-1232. DOI: <https://doi.org/10.1016/j.jare.2013.12.005>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2090123213001495>.
- [140] S. R. Islam, D. Sutanto, and K. M. Muttaqi, "A decentralized multi-agent based voltage control for catastrophic disturbances in a power system," in *2013 IEEE Industry Applications Society Annual Meeting*, 2013, pp. 1–8. DOI: 10.1109/IAS.2013.6682517.
- [141] S. Satsangi, A. Saini, and A. Saraswat, "Clustering based voltage control areas for localized reactive power management in deregulated power system," 2012.
- [142] A. Onwuachumba, M. Musavi, and P. Lerley, "Identification of critical locations of power systems," in *2017 Ninth Annual IEEE Green Technologies Conference (GreenTech)*, 2017, pp. 290–296. DOI: 10.1109/GreenTech.2017.48.
- [143] P. Song, Z. Xu, C. Luo, H. Cai, and Z. Xie, "Voltage sensitivity analysis based bus voltage regulation in transmission systems with upfc series converter," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, 2017, pp. 483–488. DOI: 10.1109/IECON.2017.8216085.
- [144] 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, ISSN: 1949-3053. DOI: 10.1109/TSG.2012.2224391.
- [145] M. L. Tuballa and M. L. Abundo, "A review of the development of smart grid technologies," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710–725, 2016, ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2016.01.011>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032116000393>.
- [146] Y. Yoldaş, A. Önen, S. Muyeen, A. V. Vasilakos, and İrfan Alan, "Enhancing smart grid with microgrids: Challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 205–214, 2017, ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2017.01.064>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032117300746>.
- [147] 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)*, 2016, pp. 659–664. DOI: 10.1109/ISGT-Asia.2016.7796463.
- [148] V. Africa, *Nigeria: Total blackout as power grid collapses*, Available at: <https://alternativeafrica.com/2018/06/16/nigeria-total-blackout-as-power-grid-collapses>, 2018 (accessed November 28, 2018).

- [149] Rayne, *Ukraine's power system hacking: Coordinated in more than one way?* Available at: <https://www.emptywheel.net/2016/01/11/ukraines-system-company-hacking-coordinated-in-more-than-one-way/>, 2016 (accessed November 28, 2018).
- [150] P. Fairley, *Averting the blackout of the century*, Available at: <http://discovermagazine.com/2016/march/15-blackout-of-the-century/>, 2016 (accessed November 28, 2018).
- [151] T. Armerding, *Cyber pearl harbor' unlikely, but critical infrastructure needs major upgrade*, Available at: <https://www.forbes.com/sites/taylorarmerding/2018/10/23/cyber-pearl-harbor-unlikely-but-critical-infrastructure-needs-major-upgrade/#6e0c4d6df8b6>, 2018 (accessed November 2, 2018).
- [152] K. Wang, M. Du, D. Yang, C. Zhu, J. Shen, and Y. Zhang, "Game-theory-based active defense for intrusion detection in cyber-physical embedded systems," *ACM Trans. Embed. Comput. Syst.*, vol. 16, no. 1, 18:1–18:21, Oct. 2016, ISSN: 1539-9087. DOI: 10.1145/2886100. [Online]. Available: <http://doi.acm.org/10.1145/2886100>.
- [153] T. AlSkaif, M. G. Zapata, B. Bellalta, and A. Nilsson, "A distributed power sharing framework among households in microgrids: A repeated game approach," *Computing*, vol. 99, no. 1, pp. 23–37, 2017, ISSN: 1436-5057. DOI: 10.1007/s00607-016-0504-y. [Online]. Available: <https://doi.org/10.1007/s00607-016-0504-y>.
- [154] L. Song, Y. Xiao, and M. van der Schaar, "Demand side management in smart grids using a repeated game framework," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 7, pp. 1412–1424, 2014, ISSN: 0733-8716. DOI: 10.1109/JSAC.2014.2332119.
- [155] D. B. Smith, M. Portmann, W. L. Tan, and W. Tushar, "Multi-source–destination distributed wireless networks: Pareto-efficient dynamic power control game with rapid convergence," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2744–2754, 2014, ISSN: 0018-9545. DOI: 10.1109/TVT.2013.2294019.
- [156] V. S. Varma, M. Mhiri, M. L. Treust, S. Lasaulce, and A. Samet, "On the benefits of repeated game models for green cross-layer power control in small cells," in *2013 First International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2013, pp. 137–141. DOI: 10.1109/BlackSeaCom.2013.6623397.

- [157] K. Wang, M. Du, D. Yang, C. Zhu, J. Shen, and Y. Zhang, "Game-theory-based active defense for intrusion detection in cyber-physical embedded systems," *ACM Trans. Embed. Comput. Syst.*, vol. 16, no. 1, 18:1–18:21, Oct. 2016, ISSN: 1539-9087. DOI: 10.1145/2886100. [Online]. Available: <http://doi.acm.org/10.1145/2886100>.
- [158] E. Yang and D. Gu, "Multiagent reinforcement learning for multi-robot systems : A survey," 2004.
- [159] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proceedings of the Eleventh International Conference on International Conference on Machine Learning*, ser. ICML'94, New Brunswick, NJ, USA: Morgan Kaufmann Publishers Inc., 1994, pp. 157–163, ISBN: 1-55860-335-2. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3091574.3091594>.
- [160] M. L. Littman, "Friend-or-foe q-learning in general-sum games," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 322–328, ISBN: 1-55860-778-1. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645530.655661>.
- [161] M. Bowling and M. Veloso, "An analysis of stochastic game theory for multiagent reinforcement learning," Aug. 2001.
- [162] A. Nowé, P. Vrancx, and Y.-M. De Hauwere, "Game theory and multi-agent reinforcement learning," in *Reinforcement Learning: State-of-the-Art*, M. Wiering and M. van Otterlo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 441–470, ISBN: 978-3-642-27645-3. DOI: 10.1007/978-3-642-27645-3\_14. [Online]. Available: [https://doi.org/10.1007/978-3-642-27645-3\\_14](https://doi.org/10.1007/978-3-642-27645-3_14).
- [163] J. Hu and M. P. Wellman, "Multiagent reinforcement learning: Theoretical framework and an algorithm," in *Proceedings of the Fifteenth International Conference on Machine Learning*, ser. ICML '98, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 242–250, ISBN: 1-55860-556-8. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645527.657296>.
- [164] K. Chatterjee, R. Majumdar, and M. Jurdziński, "On nash equilibria in stochastic games," in *Computer Science Logic*, J. Marcinkowski and A. Tarlecki, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 26–40, ISBN: 978-3-540-30124-0.
- [165] M. Feng and H. Xu, "Deep reinforcement learning based optimal defense for cyber-physical system in presence of unknown cyber-attack," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–8. DOI: 10.1109/SSCI.2017.8285298.

- [166] S. Chaitusaney and A. Yokoyama, *Influence and prevention of voltage violation and momentary electricity interruption resulting from renewable energy sources*, Aug. 2007. DOI: 10.1109/PCT.2007.4538357.
- [167] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Perolat, D. Silver, and T. Graepel, “A unified game-theoretic approach to multiagent reinforcement learning,” in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 4190–4203. [Online]. Available: <http://papers.nips.cc/paper/7007-a-unified-game-theoretic-approach-to-multiagent-reinforcement-learning.pdf>.
- [168] H. Tembine, “Deep learning meets game theory: Bregman-based algorithms for interactive deep generative adversarial networks,” *IEEE Transactions on Cybernetics*, pp. 1–14, 2019, ISSN: 2168-2267. DOI: 10.1109/TCYB.2018.2886238.
- [169] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, “A secure mobile crowdsensing game with deep reinforcement learning,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 35–47, 2018, ISSN: 1556-6013. DOI: 10.1109/TIFS.2017.2737968.
- [170] L. Yu, Y. Wu, R. Singh, L. Joppa, and F. Fang, *Deep reinforcement learning for green security game with online information*, 2018. [Online]. Available: <https://www.aaai.org/ocs/index.php/WS/AAAIW18/paper/view/17362/15586>.
- [171] M. Lai, “Giraffe: Using deep reinforcement learning to play chess,” *CoRR*, vol. abs/1509.01549, 2015. arXiv: 1509.01549. [Online]. Available: <http://arxiv.org/abs/1509.01549>.
- [172] M. Kesti, “Architecture of management game for reinforced deep learning,” in *Intelligent Systems and Applications*, K. Arai, S. Kapoor, and R. Bhatia, Eds., Cham: Springer International Publishing, 2019, pp. 48–61, ISBN: 978-3-030-01054-6.
- [173] X. Guo, S. Singh, H. Lee, R. L. Lewis, and X. Wang, “Deep learning for real-time atari game play using offline monte-carlo tree search planning,” in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2014, pp. 3338–3346. [Online]. Available: <http://papers.nips.cc/paper/5421-deep-learning-for-real-time-atari-game-play-using-offline-monte-carlo-tree-search-planning.pdf>.

- [174] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. P. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of go with deep neural networks and tree search,” *Nature*, vol. 529, pp. 484–489, 2016.
- [175] J. Heinrich and D. Silver, “Deep reinforcement learning from self-play in imperfect-information games,” *CoRR*, vol. abs/1603.01121, 2016. arXiv: 1603.01121. [Online]. Available: <http://arxiv.org/abs/1603.01121>.
- [176] C. Maddison, A. Huang, I. Sutskever, and D. Silver, “Move evaluation in go using deep convolutional neural networks,” Dec. 2014.
- [177] C. Clark and A. Storkey, “Training deep convolutional neural networks to play go,” in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML’15, Lille, France: JMLR.org, 2015, pp. 1766–1774. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3045118.3045306>.
- [178] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, pp. 354–359, Oct. 2017. DOI: 10.1038/nature24270.
- [179] A. Chaudhuri and S. K. Ghosh, “Reducing hierarchical deep learning networks as game playing artefact using regret matching,” in *Handbook of Deep Learning Applications*, V. E. Balas, S. S. Roy, D. Sharma, and P. Samui, Eds. Cham: Springer International Publishing, 2019, pp. 345–362, ISBN: 978-3-030-11479-4. DOI: 10.1007/978-3-030-11479-4\_16. [Online]. Available: [https://doi.org/10.1007/978-3-030-11479-4\\_16](https://doi.org/10.1007/978-3-030-11479-4_16).
- [180] H. Zhang, D. Li, and Y. He, “Multi-robot cooperation strategy in game environment using deep reinforcement learning,” in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018, pp. 886–891. DOI: 10.1109/ROBIO.2018.8665165.
- [181] Y. He, G. J. Mendis, and J. Wei, “Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism,” *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2505–2516, 2017, ISSN: 1949-3053. DOI: 10.1109/TSG.2017.2703842.
- [182] Z. Zhou, F. Xiong, B. Huang, C. Xu, R. Jiao, B. Liao, Z. Yin, and J. Li, “Game-theoretical energy management for energy internet with big data-based renewable power forecasting,” *IEEE Access*, vol. 5, pp. 5731–5746, 2017, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2658952.

- [183] A. Ferdowsi, U. Challita, W. Saad, and N. B. Mandayam, “Robust deep reinforcement learning for security and safety in autonomous vehicle systems,” *CoRR*, vol. abs/1805.00983, 2018. arXiv: 1805.00983. [Online]. Available: <http://arxiv.org/abs/1805.00983>.
- [184] X. Zhang, T. Bao, T. Yu, B. Yang, and C. Han, “Deep transfer q-learning with virtual leader-follower for supply-demand stackelberg game of smart grid,” *Energy*, vol. 133, pp. 348–365, 2017, ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2017.05.114>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S036054421730871X>.
- [185] Z. Ni and S. Paul, “A multistage game in smart grid security: A reinforcement learning solution,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2684–2695, 2019, ISSN: 2162-237X. DOI: 10.1109/TNNLS.2018.2885530.
- [186] L.-J. Lin, “Self-improving reactive agents based on reinforcement learning, planning and teaching,” *Machine Learning*, vol. 8, no. 3, pp. 293–321, 1992.
- [187] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller, “Playing atari with deep reinforcement learning,” *CoRR*, vol. abs/1312.5602, 2013. arXiv: 1312.5602. [Online]. Available: <http://arxiv.org/abs/1312.5602>.
- [188] R. Gylberth, R. Adnan, S. Yazid, and T. Basaruddin, “Differentially private optimization algorithms for deep neural networks,” in *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2017, pp. 387–394.
- [189] O. Wichrowska, N. Maheswaranathan, M. W. Hoffman, S. G. Colmenarejo, M. Denil, N. de Freitas, and J. Sohl-Dickstein, “Learned optimizers that scale and generalize,” in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017.
- [190] C. Gulcehre, M. Moczulski, M. Denil, and Y. Bengio, “Noisy activation function,” in *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48, New York, NY, USA: JMLR: WCP, 2016.
- [191] B. Ding, H. Qian, and J. Zhou, “Activation functions and their characteristics in deep neural networks,” in *2018 Chinese Control And Decision Conference (CCDC)*, 2018, pp. 1836–1841.
- [192] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML’10, Haifa, Israel: Omnipress, 2010, pp. 807–814, ISBN: 978-1-60558-907-7. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3104322.3104425>.

- [193] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, vol. 9, Chia Laguna Resort, Sardinia, Italy, May 2010, pp. 249–256.
- [194] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

## APPENDIX

The published research articles from the outcomes of this dissertation are given below:

- J2. S. Paul, Z. Ni, “A Learning Based Solution for an Adversarial Repeated Game in Cyber-Physical Power Systems”, Accepted for publication in IEEE Trans. on Neural Networks and Learning Systems (TNNLS).
  
- J1. Z. Ni and S. Paul, “A Multistage Game in Smart Grid Security: A Reinforcement Learning Solution,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 9, pp. 2684-2695, Sept. 2019. doi: 10.1109/TNNLS.2018.2885530.
  
- C6. S. Paul, Z. Ni, and F. Ding, “An Analysis of Post Attack Impacts and Effects of Learning Parameters on Vulnerability Assessment of Power Grid ”, [Accepted for Publication at ISGT NA 2020].
  
- C5. S. Paul and Z. Ni, “Study of Learning of Power Grid Defense Strategy in Adversarial Stage Game,” 2019 IEEE International Electro/Information Technology (EIT) Conference, Brookings, SD, pp. 1-6, May 20-22, 2019.
  
- C4. S. Paul, and Z. Ni, “A Strategic Analysis of Attacker-Defender Repeated Game in Smart Grid Security,” in Proc. of IEEE PES Innovative Smart Grid Technologies (ISGT’19) Conference, Washington D.C., pp. 1-5, Feb. 17-20, 2019.
  
- C3. S. Paul and Z. Ni, “A Study of Linear Programming and Reinforcement Learning for One-Shot Game in Smart Grid Security,” in Proc. of IEEE Intl. Joint Conference of Neural Network (IJCNN’18), Rio de Janeiro, Brazil, pp. 1-8, Jul. 8-13, 2018.

- C2. Z. Ni, S. Paul, X. Zhong and Q. Wei, “A reinforcement learning approach for sequential decision-making process of attacks in smart grid,” 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 2017, pp. 1-8.
- C1. S. Paul and Z. Ni, “Vulnerability analysis for simultaneous attack in smart grid security,” 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, 2017, pp. 1-5.