

Marquette University

e-Publications@Marquette

Dissertations (1934 -)

Dissertations, Theses, and Professional
Projects

Probabilistic Framework for Balancing Smart Grid's Performance Enhancement and Resilience to Cyber Threat

Rezoan Ahmed Shuvro
Marquette University

Follow this and additional works at: https://epublications.marquette.edu/dissertations_mu



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Shuvro, Rezoan Ahmed, "Probabilistic Framework for Balancing Smart Grid's Performance Enhancement and Resilience to Cyber Threat" (2020). *Dissertations (1934 -)*. 922.

https://epublications.marquette.edu/dissertations_mu/922

Probabilistic Framework For Balancing Smart Grid's Performance Enhancement
And Resilience To Cyber Threat

by
Rezoan Ahmed Shuvro

B.Sc., Electronics & Communication Engineering, KUET, BD, 2010
M.S., Electrical & Computer Engineering, University of New Mexico, USA, 2018

A Dissertation is Submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin
May, 2020

ABSTRACT

Probabilistic Framework For Balancing Smart Grid's Performance Enhancement And Resilience To Cyber Threat

Rezoan Ahmed Shuvro, B.S., M.S.

Marquette University, 2020

Critical infrastructures such as smart grids rely heavily on the seamless interaction between the grid subcomponents, i.e., the communication networks which transfers information from and to the grid, and the human operators/AI agents for taking necessary control actions. Smart grids are prone to cascading failures, which trigger from a few initial the tripping of a few transmission lines or generators, creating a ripple effect in the entire network, which may, in turn, lead to a total blackout. Having additional information through the communication network increases the probability of taking better control actions (e.g., effective load shedding and other protection mechanisms), which increases the reliability of the grid. On the other hand, enhancing the smart grid's communication capability increases the risk of harm through cyberattack and other faults in the communication network. A fundamental question is how can we balance the trade-off between grid's performance enhancement and robustness to information infidelity? In this dissertation, we develop a predictive analytic, scalable and tractable Markov-chain model for cascading failures in smart grids including the role of the human operators, while taking into account the benefits and harm of the communication network (e.g., supervisory control and data acquisition). The state transition probabilities of the Markov chain captures the benefits and added vulnerabilities resulting from the communication network. A detailed mapping between power-grid states and the operators' response has been established that allows capturing a wide range of operator behavior and their probabilities into in the dynamics of the Markov chain. The model shows the existence of a point of diminishing returns beyond which the harm of cyber threat and human errors outweighs the benefits of having information. An optimal level of inter-connectivity is achieved between the power grid and the communication network minimizing the expected value of the transmission-line failures.

Acknowledgement

Rezoan Ahmed Shuvro

I would like to express my deepest gratitude to my advisor, Professor Ma-jeed M. Hayat, for his excellent guidance and continuous support throughout my graduate studies. I had almost zero research experience, and he supported all aspects, including generating ideas, formulating problems, writing, and presenting research outcomes, to name a few. I have learned a lot from him, and I am grateful to him for his valuable effort and for giving me this life-changing opportunity.

I would like to thank the committee members Dr. Edwin Yaz, Dr. Ayman El-Refaie, and Dr. Richard Povinelli, for their time and suggestions. I would like to thank Dr. Zyuoyao Wang and Dr. Mahshid Rahnamay-Naeini for their guid-ance, invaluable discussions, and constructive comments on my works. My special thanks to my friend, Mitun Talukder, and Dr. Pankaz Das, for their tremendous encouragement and suggestions for the last four years. Without their support, the journey would have been a very difficult one. I would like to acknowledge the helpful faculty and staff at Marquette University and The University of New Mexico. This work was supported by the Defense Threat Reduction Agency's Basic Research Program and the National Science Foundation. Without financial support, I could not have done this work.

Finally, I thank my parents and my family for their endless support. I thank my father and mother, who never hesitated to say yes to my decision for Ph.D. studies abroad. They had to go through a lot of obstacles back at home, but always wanted to see me happy and encouraged constantly. I would like to thank my beloved wife and daughter for the amount of struggles and sacrifices they needed to do, the adjustments they needed to make to their personal life for this day to happen, which is beyond imagination and can't be expressed in words.

Executive Summary

Smart grids rely extensively on communication and sensing to provide informational data so that it operates reliably. As such, the reliability of the smart grid is heavily dependent on the interdependence between power grid components and its associated communication and control networks. The interdependency with a communication network is an inherent attribute of smart grids, as they require seamless integration of traditional power-plants, distributed energy resources (DERs) such as wind farms, solar, and micro-grids, as well as operation and management systems. For example, degradation in the communication and control networks, resulting from a cyberattack, will impact the ability to prevent or slow-down successive grid-component failures (such as transmission line failures), namely, cascading failures in the power grid which often results in blackouts. Conversely, power outages can lead to degradation in the communication and control networks, which, in turn, can further exacerbate the transmission-line and generator failures in the power grid. Additionally, human operators play an important role in smart grid operation and reliability, and they introduce a new level of interdependency between the power grid and the communication and control networks through their actions in controlling the grid using the information provided by the communication network. The system operators' expertise in dealing with cascading failures can play a pivotal role during contingencies. For instance, as the communication and control capabilities are degraded as a result of a cyber attack, cascading failures in power grids may ensue, which altogether can create a stressful and intense environment for the operators. This, in turn, increases the probability of operator errors as they diagnose and implement corrective actions (for example, through load shedding), which will further exacerbate cascading failures in the smart grid. Hence, understanding the impact of the dependence on information on the operation of the smart grid, both negative and positive, of

information-centric smart grids is critical in predicting the reliability of the smart grid.

There have been notable efforts on modeling the cascading-failure dynamics in smart grids. Here we will mention the most relevant works. Buldyrev *et al.* was the first to introduce the detrimental effects of interdependency and reported that failures in one system could collapse the entire interdependent system and lead to a blackout. Rahnamay-Naeini and Hayat used two interdependent Markov chains (for the power grid and communication networks) to capture the harm of interdependency. In their work, state-space reduction of the complex interactions of the power grid and communication components was achieved by choosing the state variables of the Markov chain inspired by observations made through simulations. Wang *et al.* modeled the effect of human error on the reliability of smart grids using performance attributes such as time to react and stress level of operators. Korkali *et al.* showed that cascading-failure risk could be reduced by increasing infrastructure network interdependence, i.e., increasing the interdependency between power grids and communication networks. However, the role of cyber threat on the reliability of communication was not captured, i.e., the harmful effect of communication was not considered. Hence, a comprehensive model capturing the benefits and harms of communication, including the role of human error in the loop has not been studied. This is a crucial driver for this dissertation.

In this dissertation, a stochastic Markov-chain model, namely, Interdependent Stochastic Abstract State Space Evolution (I-SASE), is developed, which captures the dynamics of cascading failures in the power grid and the benefits and risks of information through the communication network. To do this, a previously developed Markov-chain model is generalized to capture the trade-off between the benefits of having a robust communication infrastructure and its vulnerability to cyberattacks. The impact of error probability of system operators' as they take

actions to mitigate cascading failures is incorporated into the model as a function of the state variables of the Markov chain. The model shows that there exists a point of diminishing return beyond which the effect of cyber threat and human errors outweighs the benefits of having more information. This is one of the critical contributions of the model. An optimal level of interdependency minimizing the expected value of transmission-line failures is achieved between the power grid and the communication network.

The benefits of the I-SASE model are threefold. First, the model captures the interdependency and dynamic interactions between the layers of the power grid, which is more realistic compared to the current literature that does not include the effect of cyber threat in an analytical model. Second, the model incorporates the benefits of having inter-connectivity with a communication network through effective implementation of load shedding. Meanwhile, the model captures the harm of having excessive information through cyber threats in addition to system operator error. Our formulation of this model, therefore, leads to finding an optimal level of inter-connectivity that maximizes the benefits rendered by information, for a given level of cyber threat and operator error. Third, the model produces the probability distribution of the size of a blackout analytically, considering the potential harm from cyber threat and operator error.

The I-SASE model is analytic, scalable, and tractable. It can be beneficial for the smart grid utilities in their design process to minimize the harm rendered by cyber threats through communication infrastructure to make the smart grid robust in the face of cyberattacks. Moreover, smart grid operators can use this model to design optimal decision strategies to mitigate cascading failures while maximizing the number of customers served. The model variables and parameters can be easily scaled to apply this model in other interdependent networks such as the transportation networks.

TABLE OF CONTENTS

Acknowledgments	i
Executive Summary	ii
LIST OF TABLES	xi
LIST OF FIGURES	xvii
Chapter1 Introduction	1
1.1 Motivation	1
1.2 Examples of large blackouts and known causes	3
1.3 Literature review	6
1.3.1 Prior works on modeling cascading failures in smart grids . .	7
1.3.2 Prior works on interdependent modeling of cascading failures in smart grids	9
1.3.3 Prior works on the role of communication topology on the reliability of power grids	10
1.3.4 Prior works on embedding human errors on cascading failure modeling	11
1.3.5 Prior works on predicting cascading failures using machine learning	12
1.3.6 Prior works on optimizing interdependent system parameters for enhancing system performance	13
1.3.7 Prior works on modeling initial disturbance leading to cascading failures	14
1.4 Review of the prior Markov-chain based models used extensively in this dissertation	15

1.4.1	Review of the Stochastic Abstract State-space Evolution (SASE) model	15
1.4.2	Review of Interdependent Markov-Chain (IDMC) Approach	17
1.4.3	Review of the human SASE (hSASE) model	18
1.5	Benefits and threats of interdependency between power grid layers	19
1.6	Contribution of this dissertation	20
1.6.1	Benefits and harm of communication	21
1.6.2	Human error as a function of grid variables	21
1.6.3	Estimating average transmission capacity loss and load loss .	22
1.6.4	Role of communication topology on the reliability of power grids	22
1.6.5	Impact of initial conditions on grid reliability	23
1.6.6	Predicting Cascading Failures using machine learning algorithms	23
1.7	Organization of this dissertation	24
1.8	Publications resulted from this dissertation	25
Chapter2 Multi-layer Markov-chain based cascading-failures model		28
2.1	Simulating cascading failures in power grids	29
2.2	Cascading failure simulation (CFS) framework	30
2.2.1	Overview of the CFS framework	30
2.2.2	Power grid operating parameters and variables for cascading failures	32
2.2.3	Optimal power flow	34
2.2.4	Algorithms for finding the island and overloaded lines	35
2.3	Influence of communication network on the reliability of the power Grid	37
2.4	Modeling the communication-power interdependency function . . .	40

2.4.1	Simulation setup	40
2.4.2	Modeling the interdependency as a function of hop distance and node degree	42
2.4.3	Simulation of the IEEE 118-bus system including commu- nication network influence for different case	45
2.4.4	Simulating the Markov chain of the proposed model	46
2.5	A Markov-chain based three-layer Model for cascading-failures in smart grids	48
2.5.1	Markov chain based cascading-failure model capturing power, communication and human factors	49
2.5.2	State transitions	51
2.6	Simulation Results	55
 Chapter3 Correlating grid-operators' performance with grid vari- ables in smart-grids during cascading failures		60
3.1	Introduction	60
3.2	Markov chain based cascading failure model including operator in the loop	62
3.2.1	State variables and transition matrix of the Markov chain	62
3.2.2	Mapping between PSFs and the Markov chain state-space	65
3.3	Results	68
3.3.1	Critical PSF combinations that lead to high HEP	68
3.3.2	Comparison of blackout size with hSASE model [1]	71
 Chapter4 Balancing Smart Grid's Performance Enhancement and Resilience to Cyber Threat		73
4.1	I-SASE Analytic Framework	74

4.1.1	Influence of the communication network on the propagation of cascading failures	74
4.1.2	Influence of the human operators' on the propagation of cascading failures	78
4.1.3	State variables of the Markov chain	79
4.1.4	Probability transition matrix	80
4.2	Results	83
4.2.1	$P_{stop}(S_i)$ and blackout size distribution for various k	83
4.2.2	Optimal power-communication interdependency	83
4.2.3	Role of operators performance on the propagation of cascading failures	86
4.3	Cyber threat and operator error aware stochastic model for cascading failures	87
4.3.1	State variables of the Markov chain	88
4.3.2	Effect of interdependency among the various layers of the power grid	89
4.3.3	Transition matrix of the Markov chain	93
4.4	Results	98
4.4.1	Optimal power-communication inter-connectivity	99
4.4.2	Comparison with similar models	100
4.4.3	System operators' role on cascading failure	101
4.5	Summary of this chapter	101

Chapter 5 Dynamics of Transmission Capacity and Load Loss during Cascading Failures in Power Grids

103

5.1	Introduction	103
5.2	cSASE model	104
5.2.1	Review of the simulation framework	104

5.2.2	State variables of the Markov chain	106
5.2.3	Formulating the transition matrix	106
5.2.4	Calculating the average transmission capacity loss in the steady state	108
5.3	Results	111
5.3.1	Expected number of transmission-line failures and average transmission capacity loss	111
5.3.2	Predicting load loss from average transmission capacity loss	111
5.3.3	Critical initial conditions for various grid sizes	113
 Chapter6 Damage of initial assets can have different consequences		115
6.1	Modeling the initial failures due to the stressor(s) and impact of the stressor(s) on cascading failures in power grid	116
6.1.1	Modeling the initial failures due to stressor(s)	117
6.1.2	Simulation framework	122
6.1.3	Impact of stressor(s) event on cascading failures	122
6.2	Impact of initial failures due to a stressor event	124
6.2.1	Impact of number of failed transmission lines and capacity of the failed transmission lines	124
6.2.2	Impact of power grid loading level and load-shedding con- straint on cascading failures in power grids	126
 Chapter7 Predicting cascading failures in power grids using ma- chine learning algorithms		128
7.1	Power grid operating parameters and model features	129
7.2	Results	133
7.2.1	Description of the data	133
7.2.2	Analysis on cascading failure prediction	134

7.2.3	Linear regression to predict the number of transmission line failures and the amount of load shed	135
7.3	Predicting cascading failures with a new dataset of modified features	136
7.3.1	Features of the new data set	136
7.3.2	Data cleaning	138
7.3.3	Statistical Analysis	139
7.3.4	Results and In-depth analysis using machine learning	143
Chapter8 Summary and future works		147
	BIBLIOGRAPHY	153

LIST OF TABLES

1.1	Top 10 power grid outages due to cascading failures in the world [2]	3
2.1	Human Operator response levels based on failure in power and communication network	51
3.1	Mapping of Operator response levels with PSF levels and grid variables (nominal PSF levels considered for the other PSFs)	67
3.2	Detailed Grid variable and HEP mapping	67
5.1	α, β, γ for different values of F_i, r, e, θ	108
7.1	Prediction error	136
7.2	Hyperparameters	145

LIST OF FIGURES

1.1	Basic structure of the electric grid [3]	1
1.2	Overview of the cascading failure dynamics in power grids.	4
1.3	Timeline of the events leading to the start of cascading failures [3].	5
1.4	Real time rate of cascading-failure for the 2003 Northeast blackout [4].	6
1.5	Real-time rate of cascading failure for the 1996 blackout [5].	7
1.6	The analytical and empirical conditional PMF of the blackout size (a) without stress (b) with stress for $F_i=2$ and $C_i^{max}=20\text{MW}$ [5]. .	17
1.7	Probability of reaching a blackout follows historical blackout trends for (a) with $F_i=3$ and (b) with $F_i=6$, and different values of C_i^{max} [5].	17
2.1	IEEE 118-bus topology (control-center marked). Green and orange mark defines the clusters we used for simulation in Section 2.4.3 [6].	29
2.2	Flowchart of the cascading failure process in power grid	31
2.3	Empirical probability distribution, over the number of failed com- munication links, of the minimum hop distance of the center node to failed links (left) and the maximum degree (right) of failed nodes.	40
2.4	Empirical probability distribution of the communication-link failure over the minimum hop distance of the failed links (left) and the maximum node degree of the failed nodes (right).	41
2.5	Hop distance (left) and node degree (right) distribution as a func- tion of failed communication links.	43
2.6	Blackout size distribution, ten initial random failures each with certain hop distance of two, four and six.	45
2.7	Blackout size distribution for the two cluster-failure scenarios.	46

2.8	Blackout distribution (log-log scale) obtained from simulating the proposed model and the coupled power-to-communication simulator.	47
2.9	State transition diagram for the three layer Markov chain	50
2.10	Cascade-stop probability with and without the influence of communication and human response.	55
2.11	Blackout size (log-log scale) with and without the impact of communication and human operator influence.	56
2.12	Blackout size (log-log scale) comparison when the grid is not in stress ($r = 0.5$, $e = 0.1$ and $\theta = 0.1$ was considered for no stress scenario).	57
2.13	Blackout size (log-log scale) comparison when the grid is under stress. ($r = 0.8$, $e = 0.2$ and $\theta = 0.2$ was considered for stressed scenario).	58
3.1	HEP distribution and joint probability of PSFs	68
3.2	Comparison of blackout size between the proposed model and hSASE model [4].	70
4.1	An interdependent smart grid that integrates the power-grid, the communication network (with cyber attackers), and the grid operators.	74
4.2	A schematic diagram showing the inter-connectivity between the power grid and the communication network in smart grids	75
4.3	(a) Changes in load shedding constraint and (b) the probability of cyber threat due to power failures and $k=0.6$	78

4.4	Cascade stop probability for $k = 0.5$ and 0.9 . For $k = 0.5$, P_{stop} is high initially compared to $k = 0.9$, which indicates that the probability of a large cascade for $k = 0.5$ is lower. To visualize the effect of cascade stop probability, the y -axis is truncated from $[0,1]$ to $[0,0.2]$	84
4.5	Distribution of the blackout size for $k = 0.1$, $k = 0.5$ and $k = 0.9$. For $k = 0.1$ and $k = 0.9$, blackout size distribution shows heavy tail, which indicates a power law distribution as compared to an exponential distribution for $k = 0.5$	84
4.6	Optimal power-communication interdependency with and without cyber threat. Without cyber threat, $E[F_i S_0]$ increases with k with minimum at $k = 1$. Including the cyber-security parameter, ψ_i , $E[F_i S_0]$ is minimum when $k = 0.5$	85
4.7	Comparison of the effect of various levels of power-communication interdependency reported in [7] and I-SASE model (a) % of served loads after cascade ends for various interdependency levels (adopted from [7]) (b) % of survived lines for various interdependency levels using I-SASE model	87
4.8	Distribution of the blackout size for Nominal and extreme HEP and $k=0.5$, $F_i=2$. $E[F_i S_0]$ for nominal and extreme HEPs are 24 and 32 lines respectively.	88
4.9	A flowchart for mapping HEP with the state variables of the Markov chain	91
4.10	Human error probability mapped with state indexes of the Markov chain for various operators' PSFs	93
4.11	Schematic diagram of the state transitions of the Markov chain . . .	94

4.12	Expected transmission line failures for various inter-connectivity, and cyber threat level. We considered $F_i=2$, $C^{max}=20$, $r =0.7$, $e=0.1$ and nominal human error level. We can observe a point of diminishing return when cyber threat is considered.	99
4.13	Distribution of the blackout size given initial condition of the grid for various nominal and extreme system operator error scenario,. We considered $F_i=2$, $C_i^{max}=20$, $\Psi=1$, $r =0.7$, $e=0.1$	100
5.1	Expected number of line failure and expected transmission capacity loss for different total number of transmission lines M and initially failed transmission lines F . Here, C_i is the capacity of the latest transmission line failed among F transmission lines and $r = 0.85$, $e = 0.45$, $\theta = 0.2$	110
5.2	Predicting average load shed from cumulative transmission capacity of the failed transmission lines.	113
5.3	Expected transmission capacity loss during cascading failures for different M , F (red indicates a cascade). Here $C_i = 20MW$ i, $r = 0.85$, $e = 0.45$, and $\theta = 0.2$	113
5.4	$P(F_i \geq 40)$ and ATCL for various r , $e = 0.45$, $\theta = 0.2$. For $r \geq 0.65$, probability of a cascade increases with r	114
6.1	Gaussian, circular and linear attacks mapped into a 2-D topological space. Failures in the power grid depends on the intensity of the stressor(s). Initial transmission line failures in the power grid calculated after an attack is simulated [8].	116
6.2	IEEE 118-bus and 300-bus topology	118

6.3	Average number of failed transmission lines in IEEE 118-bus topology due to Gaussian, circular and linear stressor(s) with various intensities.	120
6.4	Average number of failed transmission lines in IEEE 300-bus topology due to Gaussian, circular and linear stressor(s) with various intensities.	120
6.5	Number of failed transmission lines when one stressor location with multiple failures (blue) and considering randomly distributed failed transmission lines (green) where a stressor event contribute one transmission- line failure (we pick the line with maximum intensity to fail).	124
6.6	Number of cascading failure event in a power grid with different number of attack points and number of transmission line failures. .	125
6.7	Relationship between the number of initially failed transmission line due to a stressor event with a percentage of additionally failed lines due to cascading when the total capacity of the failed transmission lines are fixed, and the total capacity of the initially failed transmission lines with additional capacity lost due to cascading when the number of the failed transmission lines are fixed.	126
6.8	Dependence of the average number of failed transmission lines on power grid operating parameters (l and θ).	127
7.1	Correlation among the features	130
7.2	Comparison of the overall accuracies for predicting cascading failure in power grids using machine learning algorithms	133
7.3	Comparison of the accuracies for predicting cascading failure in power grids using machine learning algorithms for the three classes of cascades.	134

7.4	Predicting the number of line failures using Linear Regression . . .	135
7.5	Predicting the cumulative amount of load shed using Linear Regression	135
7.6	box plot of the output variables with different r	137
7.7	Number of islands formed after the initial failures	138
7.8	Histogram of the failed lines due to cascade	139
7.9	Histogram of the load-shedding	140
7.10	Cascading effect vs installed capacity	141
7.11	Generation and served load vs load shedding, demand vs r	141
7.12	Correlation between features for the new dataset	142
7.13	Histogram of the cascading effect	142
7.14	Steps used for modeling	144
7.15	Classification accuracy	145
7.16	Classification accuracy for No cascade/cascade prediction	145
7.17	Regression on cascading effect	146
7.18	Prediction and test data	146

Chapter 1

Introduction

1.1 Motivation

The North-American electric grid is one of the greatest engineering marvels of the current age and is one of the biggest connected network consisting of three sub-units: the generating units, the transmission network, and the distribution units. A basic structure of the electric grid components can be visualized from Figure 1.1. Transmission lines distribute power from generating units to the customers. The operation of modern power grids (i.e., smart grids) is a prime example of a coupled interdependent system, with highly interdependent subsystems: the power-grid, the communication network, operation, and the management systems. Grid operation relies heavily on the communication network for monitoring and control, while human operators remain key elements in reliable operations of these systems [9]. The technological advancements in computer-based communication networks made it indispensable to use the developments in communication to make the power grid more intelligent. Communication networks in smart grids play an important role in the reliability of power delivery. Due to the complex nature of the operation, power grids are prone to large outages initiated from small disturbances referred to as cascading failures [10] that often lead to a partial or complete

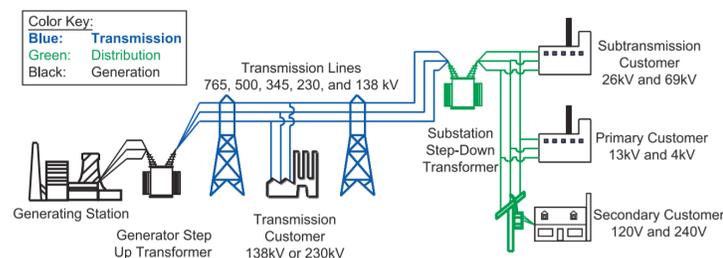


Figure 1.1: Basic structure of the electric grid [3]

blackout of the grid. Cascading failure in the power grid is defined as a sequence of correlated failures of individual components that successively weakens the power system [11]. Cascading failures can be triggered by a wide range of events, including natural disasters, technical error, human error, and intentional sabotage attacks [12]. The triggers to massive blackouts (due to cascading failures in the transmission grid) are many, including natural disturbances (ice storms, hurricanes, tornadoes, and earthquakes) as well as non-natural events such as human errors, equipment failures, cyber-attacks, weapons of mass destructions (WMDs), High altitude Electro-Magnetic Pulses (HEMP), sabotages, and supply shortages [13, 12]. When the grid is stressed as a result of failures due to natural disasters and attacks, successive failures (in power grid and communication network) may propagate within and across these networks leading to large-scale cascading-failures and blackouts. Failures in the power grid can affect the communication network, which in turn, may affect the power grid, and so on. Further, failures in the power grid can lead to failures in the communication and control network and create a stressful environment (with the high possibility of human errors) to react to the situation and implement corrective actions, which can exacerbate failures in the grid. Besides, as smart grids move toward becoming progressively more distributed and information-centric, the concern over cyber-security threats becomes increasingly alarming. With the advent of smart grids and the integration of complicated communication networks, the number of massive blackouts is occurring more frequently than before. A list of top ten power grid outages due to cascading failures are shown in Table 1.1. The number of people getting affected by these events and the economic loss is astronomical. From 1965 to 2008, nine massive blackout events were affecting more than 20 million people, whereas, in the last decade, there were seven massive blackouts, including the largest one in India [2].

Table 1.1: Top 10 power grid outages due to cascading failures in the world [2]

Location	People affected (millions)	Year
India	620	2012
India	230	2001
Bangladesh	150	2014
Pakistan	140	2015
Indonesia	2019	120
Indonesia	2009	100
Brazil	1999	97
Brazil, Paraguay	60	2009
Italy, Switzerland	56	2003
United States, Canada	55	2003

An overview of the cascading failures was depicted in Figure 1.2. In summary, the Figure shows that there is a list of power grid operating parameters that control the grid behavior. On top of that, there are interdependencies between the power grid layers, such as power-communication interdependency, power-human interdependency, and so forth. With an initial trigger (for example, transmission-line failures) resulting from a natural disaster or intentional/unintentional initial event, there can be a series of events leading to cascading failures. Human operators are an essential component of the smart-grid, and they introduce a new level of interdependency between the power grid and the communication and control networks. Due to the complexity of the network, statistical measures such as the distribution of line failures are extremely complicated to find, and in this dissertation our prime goal would be to model the interdependencies in smart grids using a Markov-chain based model and then predict the nature of cascading failures in the power grid from the initial conditions of the grid during cascade initiation.

1.2 Examples of large blackouts and known causes

Historical data of the power grid failures suggest that various factors related to these elements can affect the efficient operation of the power grids and contribute to cascading-failure. Various factors related to the elements of the smart

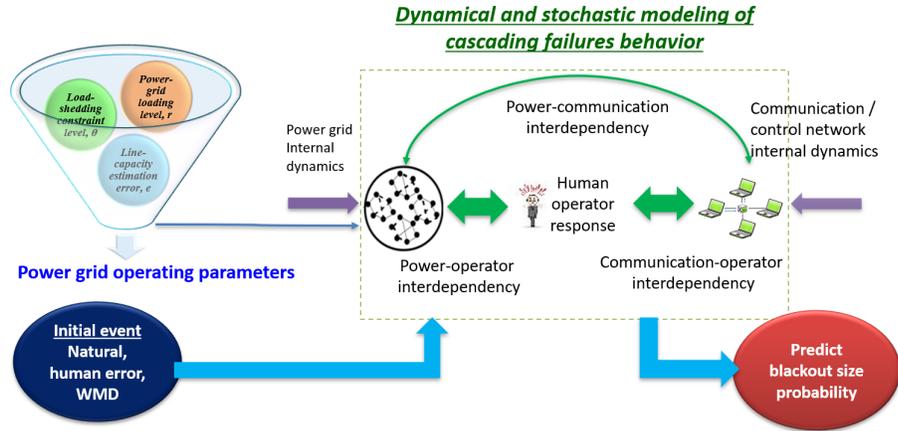


Figure 1.2: Overview of the cascading failure dynamics in power grids.

grid can affect the efficient operation of the grid and contribute to cascading failures. For example, the 2003 Northeast blackout in the United States and Canada occurred due to a combination of transmission-line and generating-unit failures, communication component and server failures, as well as ineffective and erroneous human-operator responses as observed in Figure 1.3 taken from the postmortem report by the North American Electric Reliability Corporation (NERC) on 2003 cascading failures event [3]. Specifically, the alarm software failed, leaving the human operators unaware of the transmission-line outage, which contributed to the cascading-failure [14]. After the initial trigger of failure in the transmission-line due to a combined impact of power grid and communication system disturbances, cascading-failure evolved in phases throughout the power grid network (as seen from Figure 1.4) and resulted in more than 55 GW's of power failure [4]. Figure 1.4 shows the three phases of the cascading failures namely, precursor phase, escalation phase and the cascade phase-out phase [1]. Similar evolution of cascading failure phases was also observed in July 1996 and August 1996 in the Western Interconnection [5] as shown in Figure 1.5.

The power blackout in Italy of 2003 is another example of power grid and communication network interdependency where an unplanned power shutdown

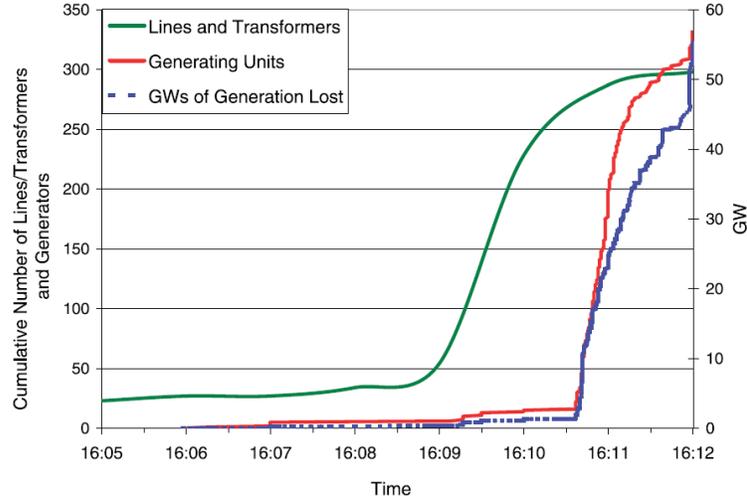


Figure 1.4: Real time rate of cascading-failure for the 2003 Northeast blackout [4].

to study the interdependencies among these coupled systems to understand the role of such interdependencies so that cascading-failures can be prevented before massive blackouts.

1.3 Literature review

In this section, we review the works related to this dissertation. Besides, we also point out the limitations of these works and briefly describe the contributions of this dissertation where necessary. There have been extensive works on modeling cascading failure in the power grid. Efforts can be categorized in mainly in three approaches: (i) network/graph-theoretic approaches (including complex-system theory) [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28] , (ii) power-system simulations [29, 30, 31, 32, 33, 34, 35, 23, 36, 37, 38, 39, 40, 30, 41, 42, 43] (iii) probabilistic analytical models [29, 31, 7, 5, 44, 45, 46, 47, 41, 45, 48, 49, 50, 51, 52, 53]. Probabilistic analytical models usually use concepts from branching processes, regeneration theory, percolation theory, or Markov chains to model the stochastic-failure dynamics in the power grid. Probabilistic models on cascading failures [13] in the power grids focus on both failure dynamics in the power grid in a single, non-interacting environment or considering the interdependence between

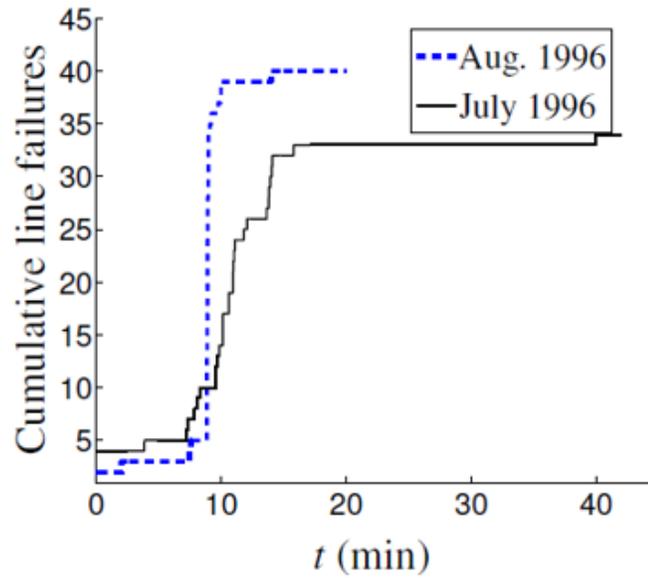


Figure 1.5: Real-time rate of cascading failure for the 1996 blackout [5].

the power grid and the communication network.

One general weakness in all the prior works is the insufficient treatment of the realistic functional interdependency among cyber threats, human factors, and the power grid. The prime focus of this study is to address both the benefits and harm of interdependence of the communication network and find a balance between the smart grid’s performance enhancement and resilience to the cyber threat.

1.3.1 Prior works on modeling cascading failures in smart grids

Cascading failure dynamics involve several power-grid variables and interdependency between power grid and communication system variables [7, 30], modeling cascading failures, and analyzing the severity of cascading failure is a challenging task. Nonetheless, since the 2003 blackout in North America [54], significant research efforts have regarded models of cascading failure dynamics in power grids and strategies to understand and mitigate the risk of cascading failures [13]. A common approach is to use of probabilistic modeling of cascading failures in the power grid either as an independent system [32, 5, 55] or in the presence of an

interdependent system environment [44, 30, 21, 7] (e.g., the interdependency between the power grid and communication system). Cascading failure analysis tools, guidelines, and metrics are reported by the Institute of Electrical and Electronics Engineers (IEEE) cascading failure working group in [11, 56, 57, 58].

A Markov-chain driven probabilistic model named stochastic abstract state evolution (SASE) was proposed in [5] that captures the dynamics of cascading failures in a power grid using a reduced state space. The notion of an equivalent class in [5] allows modeling the large state space of the power grid with infinitely many variables by using a small equivalent reduced state space. More recently, in the human-SASE (hSASE) model [1], the authors considered an additional state-variable to capture the effect of human operator errors on the power grid dynamics. Both the SASE and hSASE models are capable of determining the distribution of transmission-line failures and evaluating the blackout size at steady-state. In [59], a data-driven parametric model is proposed to characterize the dynamics of the propagation of transmission-line failures of the power grid. This model can estimate the total capacity loss due to the failure of transmission lines numerically during cascading failures at any time step. In contrast to [5], multiple failures per time unit were allowed, and a transition probability was associated with failures affecting transmission lines having lower capacities [59]. However, this data-driven model lacks a key feature of previous models as it does not include the probability of the cascade stopping at any state. Moreover, one standard limitation in most of the other cascading failure probabilistic models including [5, 1, 59] is that they can not predict transmission capacity loss/amount of load-shedding, which is considered to be a critical metric for evaluating the severity of cascading failure by the IEEE cascading failure working group [58].

1.3.2 Prior works on interdependent modeling of cascading failures in smart grids

During the initial stages of cascading-failure analysis in the power grid, most of the studies were focused on modeling and analyzing failure dynamics in the power grid in a single, non-interacting environment [11, 5]. Recently, cascading-failure studies in interdependent systems are emerging. Notably, Buldyrev *et al.* reported a graph-based interdependent network study that showed the coupling impacts between interdependent networks while analyzing the percentage of failed nodes in the steady-state [21]. In [30], the authors proposed a two-phase control policy to mitigate the cascading-failure in the power grid using the interdependency with the communication network. Carreras *et al.* coupled with two complex systems and investigated the effect of coupled system interactions between infrastructure systems. [31]. The impact of communication topology on the propagation of cascading-failure in the power grid was analyzed in [60] using graph theory. The vulnerability of the power grid in a coupled power-communication grid environment was analyzed in [7], which showed that interconnecting networks could enhance robustness. The impact of various initial failures on physical infrastructure (e.g., power, communication networks) were analyzed in [8].

Our point of interest is in the probabilistic models, which can be further categorized by the following approaches: Markov-chain based models [5, 45], branching processes [61], regeneration theory [55], etc. The Interdependent Markov Chain (IDMC) model [44] by Rahnamay-Naeini *et al.* captures cascading-failure in a interdependent system environment (e.g., smart grid). The model considered two interdependent Markov chains consolidated in a single Markov chain, whose transitions alternate between transitions in the two interdependent networks while capturing the interdependencies between the two systems represented by Markov

chains. As an example, the model considered an interdependent power grid and communication network, and the IDMC model captures the transitions among power and communication networks' state variables.

A communication-power coupling parameter d , which captures the topological changes in the communication network, was modeled in [62]. In recent work, Wang *et al.* presented an analytic framework based on a Markov chain for modeling the dynamics of infrastructure under contingencies, while capturing the effects of operators' behavior quantified by the probability of human error under various circumstances [1]. To quantify the human error during cascading-failures, the SPAR-H methodology [63] is used to estimate the human error probability (HEP) for the human performance status in a probabilistic risk assessment approach [64]. The work presented in [64] captures the coupling between the human factor and the power grid in their proposed analytic model.

1.3.3 Prior works on the role of communication topology on the reliability of power grids

Most of the studies on cascading-failures have focused on cascading-failures in a single, non-interacting power systems environment [11]. However, in recent days, a body of work is emerging in studying the cascading-failures in interdependent systems and specifically, the power and communication networks. These studies aim to identify the behaviors of interdependent systems and how they interact with each other during a failure event. The efforts can be categorized into three classes: probabilistic analytical models [52, 61, 44], deterministic analytical models [47] and analysis of failure scenarios using simulations [29]. Our point of interest is the probabilistic models, which can be further categorized by the following approaches: Markov-chain(MC) based models [5, 1], branching processes [65, 61], regeneration theory [44], etc. As mentioned previously, Buldyrev *et al.* reported a graph-based interdependent network study that showed the coupling

impacts between interdependent networks [21]. In [30], the authors proposed a two-phase control policy to mitigate the cascading-failure in the power grid using the interdependency with the communication network. Carreras *et al.* coupled two complex systems and investigated the effect of coupling on the characteristic properties of that system [31]. The impact of communication topology on the propagation of cascading-failure in the power grid was analyzed in [60]. In a recent study, Wang *et al.* showed the interdependency between the power grid and human operator response during cascading-failures [1].

1.3.4 Prior works on embedding human errors on cascading failure modeling

In recent years, a substantial amount of work has been done to understand the role of interdependencies between smart grid subsystems and the dynamics of cascading failures [21, 7, 20, 30]. Reviews of models modeling cascading failure dynamics in power grids were reported in [11, 13]. Bench-marking of quasi-steady-state cascading outage analysis methodologies were reported in [58]. However, none of these works includes the role of human error in modeling cascading failures in power grids. To quantify the human error, Standardized Plant Analysis Risk - Human (SPAR-H) [66] methodology was used to identify the critical human operator attributes, performance shaping factors, and their associated levels [63, 67].

Recently, based on grid-operator interviews, Joana *et al.* [64], proposed HEP formulation depending on the performance shaping factors (PSF's) of the grid operators using the SPAR-H methodology. The probabilities for each of the PSF levels were calculated based on the smart-grid operator interviews [64]. A notion of a comprehensive model including the power grids, communication network, and the role of human operators were introduced in [68]. However, the details of the state transition probabilities were absent.

1.3.5 Prior works on predicting cascading failures using machine learning

A proactive blackout prediction model for a smart grid early warning system was proposed in [69]. In that work, a support vector machine (SVM) has been trained with this historical database and is used to predict blackout events in advance. The critical contribution of that paper is that it captures the essence of the cascading failure using the probabilistic framework and integration of the SVM machine learning tool to build a prediction rule, which would be able to predict the scenarios of the blackout as early as possible. However, their data set includes only 50 cases, which are then used for training and testing purposes. It is challenging to learn the complex dynamics of cascading failures in the power grid using 50 test cases only. Also, the authors reported that for specific parameter values, 100% training and testing accuracy was achieved, which is very unlikely and unrealistic for a sophisticated event like cascading failures. Nonetheless, the paper is a novel work on proactive cascade prediction using a machine learning approach. In [70], the authors proposed a machine learning based on Bayes networks to predict cascading failure propagation. Their model named ITEPV collects power grid data from simulations, and then their model predicts cascading failure propagation with the highest probability using the machine learning technique. This paper is another work that focuses on data-driven cascade prediction using a machine learning approach, but the authors did not describe how they collected the data, how they simulated power flow, what simulation software they used, which make reproducibility of the work difficult. A classification problem was formulated that classifies a cyber attack from other classical disturbances in the power grid in [71]. The authors used various machine learning algorithms to evaluate classification performance and tried to find the optimal algorithms under any given

constraints. The authors observed various measures (accuracy, precision, recall, and F-Measure) to show that Adaboost+JRipper is the optimal algorithm for classifying various types of cyber-threats in the power grid. The work is an initial benchmark for disturbance classification in the power grid. The authors used the WEKA [72] machine learning framework for implementing various algorithms. Benchmarking of various deep learning algorithms and comparison of %RMSE reduction by the different algorithms from existing state-of-the-art for load forecasting in smart grid applications was done in [73]. To the best of the authors' knowledge, no work has been found to classify cascading failures in power grids as well as predicting essential attributes such as the number of failed transmission lines and the amount of load shedding for an initial disturbance conditional on the power grid operating parameters and the topological parameters of the grid. One of the main reasons is the unavailability of a real-world cascading failure data set.

1.3.6 Prior works on optimizing interdependent system parameters for enhancing system performance

Probabilistic models on cascading failures [13] in the power grids focus on both failure dynamics in the power grid in a single, non-interacting environment [11, 5, 29] or considering the interdependence between the power grid and the communication network [7, 21, 30, 44]. Two complex systems were coupled to investigate the effect of coupled system interactions between infrastructure systems in [31]. Buldyrev *et al.* reported the detrimental effects of interdependence in a graph-based study, which showed that failures in one system could collapse the entire interdependent systems and lead to a blackout [21]. A two-phase control policy to mitigate the cascading-failure in the power grid using the interdependency between the power grid and communication networks was proposed in [30]. A Markov-chain based probabilistic model was proposed in [5], which captures the dynamics of cascading failure in a power grid using a reduced state space. Sub-

sequently, the authors proposed an interdependent Markov Chain model [44] that captures interdependence; however, the model only captures the negative aspects of coupling. In contrast to [21, 44], [7] showed robustness can be enhanced by interconnecting networks. Rather than having an optimal interdependence reported in [20], the authors in [7] reported that the risk of cascading failures is minimum when the infrastructure network interdependence is maximum. Although existing models capture the influence of communication networks, few works capture the role of human error on cascading failures mathematically. In recent work, Wang *et al.* presented a Markov-chain model to capture the interdependence between human error and cascading failures in power grids [1]. However, the authors did not consider the role of communication networks. One general weakness in all the prior works is the insufficient treatment of the realistic functional interdependency among cyber threats, human factors, and the power grid. The prime focus of this study is to address both the benefits and harm of interdependence of the communication network and find a balance between the smart grid's performance enhancement and resilience to the cyber threat.

1.3.7 Prior works on modeling initial disturbance leading to cascading failures

In the last two decades, both single and interdependent models were proposed by researchers to capture the dynamics of the cascading failure in the power grid. Our focus is to study probabilistic models which can be further categorized to Markov-chain based models [45, 5, 1], branching processes [61], regeneration theory [55]. These models analyze the cascading failures in the power grid based on an initial event. The interdependent system model [44, 30, 74, 21, 7, 75, 68, 76] capture the interdependency between layers of the power grid (e.g., power grid, communication system, and human-operator response) and analyzes cascading failures in the power grid based on interdependent system environment. A data-

driven model for simulating the evolution of transmission line failure in power grids is proposed in [59]. Although failures in the communication layer and human operator responses are crucial in cascading failure analysis, we ignored their effects in this paper to simplify our analysis. Bernstein *et al.* analyzed the power grid vulnerability due to geographically correlated failures in [34]. Impacts of operating characteristics on the sensitivity of the power grids to cascading failures are studied in [77]. In [60], the authors studied the impact of topology in power grids. In [8], the authors analyze the impact of various initial failures on physical infrastructures (e.g., communication networks).

1.4 Review of the prior Markov-chain based models used extensively in this dissertation

In this section, we review the prior Markov-chain based models used extensively in this dissertation.

1.4.1 Review of the Stochastic Abstract State-space Evolution (SASE) model

There are many physical attributes of the smart grids that collectively contribute to cascading failures, including power generation, substation loads, power flow distribution through transmission lines, the functionality of grid components, voltage and phase of transmission lines and buses, and so forth. This results in substantial detailed state space. It is, therefore, essential that effective spate-space reduction is performed before any scalable and tractable analytical model of cascading failures can be developed. To address the scalability challenge, a methodology was developed in [5], where the space of all detailed power-grid states is partitioned into a collection of equivalence classes. The equivalence relation is defined through a set of reduced state variables that are deemed to govern the cascading behavior as determined by extensive analysis of experimental and sim-

ulation data. The detailed power-grid states in the same class are represented by a few aggregate state-variables, with the same values. Such partitioning of the state space (of the detailed state-space of the power grid) implies that detailed power-grid states in the same class will be indistinguishable as far as the cascading behavior is concerned. Each class of the power grid states is termed an abstract state.

Specifically, the physical variables defining each equivalence class are, number of transmission-line failures, F , the maximum capacity of the failed lines, C^{max} , and a binary variable, I , which depends implicitly (and non-linearly) on all the detailed variables. It is judiciously introduced to capture the complex event of the cascade stopping compressively. If the power grid is in a cascading mode, then $I=0$, and the cascade will continue. Conversely, the cascading failure terminates if the power grid is in an absorbing state, namely when $I=1$. To address the dynamic nature of cascades, a Markov chain is defined on the reduced state space, and the transition probabilities have been extracted using physics-based modeling combined with data analytics. Since the formation of the transition probabilities is learned from data, they capture the role of the hidden variables that have been eliminated in the state-space reduction. A key component in the transition probability matrix is the state-dependent cascading-stop probability, P_{stop} , i.e., the probability that the binary state variable $I=1$.

The SASE model developed in [5] can effectively predict the underlying distribution of the blackout can be calculated analytically using the model. The model was validated using the comparison between the distribution obtained using Monte Carlo simulation of the grid and the analytical result as shown in Figure 1.6. The simulated distribution and the analytical distribution closely match, i.e., the model is able to find the distribution of blackout size analytically using the evolution of the Markov chain which is independent of the simulated result obtained

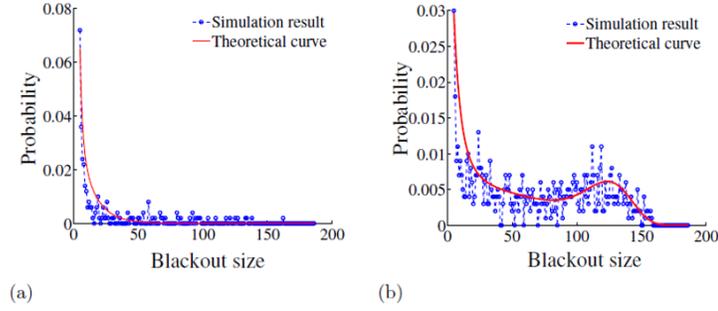


Figure 1.6: The analytical and empirical conditional PMF of the blackout size (a) without stress (b) with stress for $F_i=2$ and $C_i^{max}=20\text{MW}$ [5].

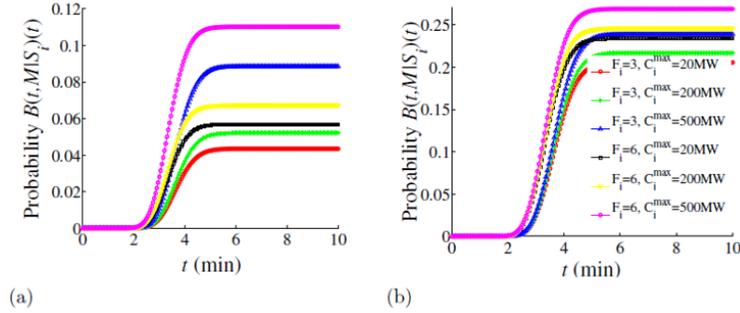


Figure 1.7: Probability of reaching a blackout follows historical blackout trends for (a) with $F_i=3$ and (b) with $F_i=6$, and different values of C_i^{max} [5].

from Monte Carlo simulations of the smart grid.

Again, the model can predict the evolution of cascading failures for various state variables and operating condition of the grid as observed in historical blackout trends in Figures 1.4 and 1.5.

1.4.2 Review of Interdependent Markov-Chain (IDMC) Approach

To model the cascading-failure dynamics, a Markov-chain based Stochastic Abstract State Evolution (SASE) model was developed in [5]. The state-space of the Markov chain consisted of two state variables: the number of failures in the system, $x_n (n \geq 1)$ and a $\{0, 1\}$ -valued variable termed cascade-stability of the system, i_n , for which $i = 0$ indicates the system is susceptible to further failures and $i = 1$ indicates that no further failures are possible, and hence the system has entered a cascade-stable mode. The SASE model allows us to predict the

distribution of the failed transmission-lines as a function of time, and also allows us to calculate the probability distribution of the blackout size. The transition probability matrix in the SASE model involves physical operating parameters of the power system, and its entries are state-dependent. The operating characteristics considered in the SASE model are based on power grid simulations, and they include line-tripping threshold e , power-grid loading level r , and the load-shedding constraint level θ .

In subsequent work, the SASE model was expanded to the IDMC model in order to capture cascading-failure in the interdependent system environment (i.e., smart grid) [44]. The model considered two interdependent Markov chains consolidated in a single Markov chain, whose transitions alternate between transitions in the two interdependent networks. As an example, the model considered power grid and communication network, and as the Markov chain progresses, transition toggles between power grid variables and communication network state variables. Specifically, the state-space of the IDMC model contains variables of the form $S_n = (x_n, i_n, y_n, l_n, k_n)$, as compared to $S_n = (x_n, i_n)$ in the SASE model. The newly-introduced variable y_n is defined as the number of failures in the communication network. Moreover, l_n captures the state transition turns between the two networks, where $l_n = 0$ indicates the last transition has occurred in the power grid, and $l_n = 1$ indicates the last transition has occurred in the communication network. Finally, the state variable K_n captures the history of transitions: it indicates whether a failure has occurred in the last transition at time $n - 1$. However, the IDMC model does not explicitly capture the role of the dynamic topological attributes of the communication network.

1.4.3 Review of the human SASE (hSASE) model

Wang *et al.* developed a Markov-chain based human-stochastic abstract state evolution (hSASE) model including the role of human error [1], where the space of

all detailed power-grid states is partitioned into a collection of equivalence classes. The model captures the cascading failures in the power grid, including HEP into the state-space of the Markov chain [1]. To calculate the HEP, the authors used the HEP formulation based on performance shaping factor (PSF) level multipliers adopted from [64]. However, [1] has the following limitations. First, the mapping of the human PSFs with the power grid variables is very coarse. Second, the distribution of the PSFs is not embedded in the dynamics of cascading failures. Third, the model only considered two PSFs that have a direct correlation with the propagation of failures (out of the eight available in [64]). Other factors, such as experience, work process, and procedures, were ignored.

1.5 Benefits and threats of interdependency between power grid layers

Unlike traditional power grids, the interdependency with a communication network is an inherent attribute of smart grids, as they require seamless integration of traditional power-plants, distributed energy resources (DERs) such as wind farms, solar, and microgrids, as well as operation and management systems. Increased interdependency between the power grid and the supporting communication networks allows power-grid operators to detect faults and implement robust control actions to enhance the resilience of the grid. On the negative side of having more interdependency, degradation in the communication and control networks, resulting from a cyber-attack, will impact the ability to prevent or slow-down cascading failures in the power grid. Higher grid-communication interdependency increases the exposure to cyber threat: hackers/attackers may intercept communication networks and alter the control decisions by altering sensor data, injecting false data, and so forth.

Conversely, power outages can lead to degradation in the information flow from power nodes to the operators at the control center. Additionally, human operators are also a critical component of the smart grid, and they introduce a

new level of interdependency between the power grid and the communication and control networks. For instance, as the communication and control capabilities are degraded as a result of a cyber-attack, cascading failures in the power grids can ensue, which altogether can create a stressful and intense environment for the operators, which, in turn, increases the probability of operator errors as they diagnose and implement corrective actions, which further exacerbates the cascade in the smart grid [13, 10].

Based on the real-world examples provided above, it is of great importance to study the interdependencies between these coupled systems (e.g., power grid and communication network) in order to understand the coupling effects so that cascading-failures can be prevented under stressed scenarios. Understanding the impacts, both negative and positive, of information-centric smart grids is critical in predicting the reliability and the resilience of the smart-grid. A robust interconnection strategy is required to balance the trade-off between grid's performance enhancement and the resilience of the grid to cyber threat, i.e., maximize the benefit of supervisory control and data acquisition (SCADA) information while containing the exposure to the cyber threat which is the main thrust of this dissertation.

1.6 Contribution of this dissertation

The main thrust of this dissertation is to analyze the strong interplay between the power grid and the corresponding communication and control network, which plays a pivotal role in the reliability and resilience of the smart grid. The dynamics of the interdependence among smart-grid subsystems such as the power grid, communication network, and response of human operators are captured during the propagation of cascading failures.

1.6.1 Benefits and harm of communication

The first contribution of this dissertation (discussed in detail in chapter 4) is that a previously developed Markov-chain based model is refined into an interdependent Markov chain model to capture the role of cyber threat from the communication network and the human-operator error during cascading failures. The state transitions of the Markov chain are parameterized by the critical operating parameters of the power grid. The calculations assume a generic form of correlation between the level of and damage from cyber-attacks, on the one hand, and the level of interdependence on the other hand. The model finds the optimal level of interdependence, i.e., the trade-off between well-informed control and vulnerability to attacks that minimizes the probability of massive cascading failures in power grids. There is a point of diminishing returns beyond which the harm of exposure to cyber threat outweighs the benefits of information.

1.6.2 Human error as a function of grid variables

The second contribution of this dissertation is that the role of human factors associated with the grid operators, e.g., human-error probability (HEP), is modeled as a function of the grid conditions as well as operators' training and experience levels (discussed in detail in chapter 3). Moreover, the HEP is embedded in a previously reported Markov-chain model that generates the probability distribution of the blackout as a function of time following a trigger. Specifically, through the HEP, the Markov-chain's transition matrix includes the dynamics of detailed smart-grid operator attributes. To derive the grid-state dependent HEP, three real-valued performance shaping factors (PSFs), representing critical human attributes of the operators, are mapped to the grid-state variables, thereby capturing the correlation between the evolution of the PSF levels and the propagation of transmission-line failures. This mapping is established based on a histogram-

equalization principle [78], which utilizes the experimentally-estimated probability distribution of the PSF levels while assuming a monotone relationship between the HEP values and number of line failures. Further, the distribution of the PSF levels was used to identify the critical combinations of PSF levels that correspond to an event with high joint probability as well as a high HEP.

1.6.3 Estimating average transmission capacity loss and load loss

The third contribution is the development of an analytical model to predict the average transmission-capacity loss and load loss during a cascading failure as a function of time and their steady-state values (discussed in detail in chapter 5). Cascading failures in the power grid are described using a Markov-chain approach, in which the state transition probabilities depend on the number and capacities of the failed lines. The transition matrix is characterized by parametrically using Monte Carlo simulations of cascading failures in the power grid. The severity of cascading failure is estimated using two metrics: the expected number of transmission-line failures and the amount of load shedding/load loss (inferred from the average transmission capacity loss) in the steady-state. These two metrics provide critical information regarding the severity of a cascading failure in a power grid (in terms of both the distribution of blackout sizes and the amounts of load shedding). One of the benefits of this model is that it enables the understanding of the effect of initial failures and the operating parameters of the power grid on cascading failures.

1.6.4 Role of communication topology on the reliability of power grids

The fourth contribution is the modeling of the impact of the power grid and communication network interdependencies on the reliability of the power grid by capturing the influence of the communication network on the power grid during the cascading phenomenon (discussed in detail in chapter 2). Two critical topological

parameters of the communication network are studied, namely the hop distance and the node degree, which are used to determine the behavior of the coupling parameter between the two systems and to quantify the influence of communication network in the power grid. Finally, numerical results have been carried out to quantify the impact of communication network failure on power grid reliability and validate our proposed model.

1.6.5 Impact of initial conditions on grid reliability

The fifth contribution of this dissertation is the analysis of cascading failures in the power grid under various initial conditions modeled analytically (discussed in detail in chapter 6). First, Gaussian, circular, and linear stressors are used as the initial events to model the probability of transmission line failure due to the stressors. Second, Monte-Carlo simulations are used to analyze the impact of cascading failures in the power grid based on the initial failure patterns. The reported results show that upon the occurrence of an initial triggering event, a combination of parameters (e.g., the number of stressors, the number of failed transmission lines in each stressor location, the capacity of the failed transmission lines, the power-grid loading level, the load-shedding constraints at the time of the stressor event) strongly influence the dynamics of cascading failures and may lead to massive blackouts.

1.6.6 Predicting Cascading Failures using machine learning algorithms

The sixth contribution of this dissertation is that we classify cascading failures in a power grid that leads to massive blackouts in power grids using machine learning algorithms (discussed in detail in chapter 7). Since real-world cascading failure data is not available, we create a synthetic cascading failure simulator framework to generate cascading-failure data for various power grid operating parameters. We include the topological parameters such as edge betweenness centrality, the

average shortest distance for various combinations of two transmission line failures in our data set. Then we apply various machine learning algorithms to classify cascading failures and compare accuracy. Further, we use regressive models to predict the number of failed transmission line and the amount of load shedding. This data-driven technique is useful to quickly predict and classify cascading failures based on the input power grid conditions, and hence power grid design engineers can use this to increase the robustness of the grid.

1.7 Organization of this dissertation

This dissertation is organized as follows. In chapter 2.1, we first describe our cascading failure simulator (CFS) framework. There is no straightforward and publicly available cascading failures analysis tool. However, a widely accepted MATPOWER [79] tool for power flow simulation are available where power flow analysis can be done using IEEE standard case studies (case studies can also be customized). In this dissertation, we have used the MATPOWER power flow tool for power flow analysis and developed our cascading failure simulator on top of it. The detailed framework of the simulator, along with the algorithms, will be discussed in chapter 2.1.

Next, in chapter 2, we model the impact of communication network failures on power grid reliability and present a naive analytic approach for modeling cascading-failures in power grids including communication and human operator impacts.

Then, in chapter 3, we establish a correlation between grid-operators' performance with cascading failures in smart-grids.

Next, In chapter 4, we develop the interdependent cascading failure model. First, we discuss the interdependencies between coupled layers. Next, we present a detailed discussion about the modeling of system operator errors and the use of the distribution of the operators' performance shaping factor distributions. Here

we map the human error levels as a function of the state variables of the Markov chain. Next, we discuss the trade-off between the grid's performance enhancement and risks of cyber threats. We then modify this approach to capture the dynamics of cyber threats using a state variable of the Markov chain.

Further, we refine the base SASE model in chapter 5 to model the dynamics of transmission capacity and load loss during cascading failures in power grids.

To this point, we have assumed that the initial conditions occur due to various natural and human-made events but did not model the initial conditions. In chapter 6, we model the impact of an initial stressor(s) on cascading failures in power grids.

Next, in chapter 7, we discuss how the cascading failure simulation framework developed in chapter 2.1 can be used to collect data, which can be used to feed machine learning models for regression and classification purposes. We use various machine learning algorithms and make a comparison of the model performance.

Finally, we conclude our dissertation with the summary and future directions in chapter 8.

1.8 Publications resulted from this dissertation

A list of our publications resulted from this dissertation, is as follows:

1. Rezoan A. Shuvro, Pankaz Das, Majeed M. Hayat. "Cyber Threat and Operator Error Aware Stochastic Model for Cascading Failures," to be submitted in IEEE Transactions on power systems, 2020.
2. Rezoan A. Shuvro, Pankaz Das, Majeed M. Hayat. "Balancing smart grids resilience from cyber threat," In 2019 IEEE resilience week symposium.
3. Rezoan A. Shuvro, Mitun Talukder, Pankaz Das, Majeed M. Hayat, Predicting Cascading Failures in Power Grids using Machine Learning Algorithms. In 2019 IEEE North American Power Symposium (NAPS'2019).

4. Rezoan A. Shuvro, Pankaz Das, Joana Abreu, Majeed M. Hayat, Correlating Grid-operators' Performance with Cascading Failures in Smart-Grids. In 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe).
5. Rezoan A. Shuvro, Pankaz Das, Mahshid Rahnamay-Naeini, Francesco Sorrentino, and Majeed M. Hayat. "On the Dynamics of Transmission Capacity and Load Loss during Cascading Failures in Power Grids." In 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe).
6. Rezoan A. Shuvro, Pankaz Das, Zhuoyao Wang, Mahshid Rahnamay-Naeini, and Majeed M. Hayat. "Impact of Initial Stressor(s) on Cascading Failures in Power Grids." In the 2018 North American Power Symposium (NAPS).
7. Rezoan A. Shuvro, Zhuoyao Wang, Pankaz Das, Mahshid R. Naeini, and Majeed M. Hayat. "Modeling cascading-failures in power grids including communication and human operator impacts." In 2017 IEEE Green Energy and Smart Systems Conference (IGESSC).
8. Rezoan A. Shuvro, Zhuoyao Wang, Pankaz Das, Mahshid R. Naeini, and Majeed M. Hayat. "Modeling impact of communication network failures on power grid reliability." In 2017 North American Power Symposium (NAPS).

A list of publications related to this dissertation co-authored by the author of this dissertation is as follows:

1. Zhuoyao Wang, Mahshid Rahnamay-Naeini, Joana M. Abreu, Rezoan A. Shuvro, Pankaz Das, Andrea A. Mammoli, Nasir Ghani, and Majeed M. Hayat. "Impacts of operators' behavior on reliability of power grids during cascading failures." IEEE Transactions on Power Systems 33, no. 6 (2018): 6013-6024.

2. Pankaz Das, Rezoan A. Shuvro, Kassie Povinelli, Mahshid Rahnamay-Naeini, Nasir Ghani, and Majeed M. Hayat. "A Theory for Optimizing Insecure Inter-Network Connections." *IEEE Systems Journal*
3. Pankaz Das, Rezoan A. Shuvro, Mahshid Rahnamay-Naeini, Nasir Ghani, and Majeed M. Hayat. "Efficient interconnectivity among networks under security constraint." In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pp. 88-93. IEEE, 2018.
4. Pankaz Das, Rezoan A. Shuvro, Zhuoyao Wang, Mahshid Rahnamay Naeini, Nasir Ghani, and Majeed M. Hayat. "Stochastic failure dynamics in communication network under the influence of power failure." In *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1-8. IEEE, 2017.
5. Pankaz Das, Rezoan A. Shuvro, Zhuoyao Wang, Mahshid Rahnamay Naeini, Nasir Ghani, and Majeed M. Hayat. "Stochastic failure dynamics in communication network under the influence of power failure." In *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1-8. IEEE, 2017.

Chapter 2

Multi-layer Markov-chain based cascading-failures model

Recent studies on interdependent systems have considered interactions either between the power grid and the communication network [7, 21, 30, 44]. There are also efforts in understanding the role of human operators' on the reliability of the power grids [1]. However, the study of these systems considering the interactions among power grid, communication network, and human operators' have not been presented heretofore.

In this chapter, we present a comprehensive three-layer model for capturing cascading-failure dynamics that take into account the interdependency among the three-layers' (i.e., power grid, communication network, and human factors) of the grid. The Markov-chain based model (SASE and IDMC; described in chapter 1) is used to capture the dynamics of cascading-failures. Precisely, cascading-failure dynamics in the power grid capture the internal dynamics of the power grid as well as performance and connectivity degradation's in the communication network and the human errors through the Markov chain abstraction. Failures of communication components can cause a delay in communication, which can affect critical control signals from supervisory control and data acquisition (SCADA). In the human-factor layer, operators' behavior will be affected by the status of both of the power system and the communication network. When failures occur in the power system, operators' may face stress and could make decisions that are not optimal. Hence, operators' may take wrong actions or no action to contingencies in the power system. Any of the sub-optimal choices made by the operators mentioned above can be considered as human error in the three-layer model.

A critical insight obtained from the proposed three-layer model to the smart

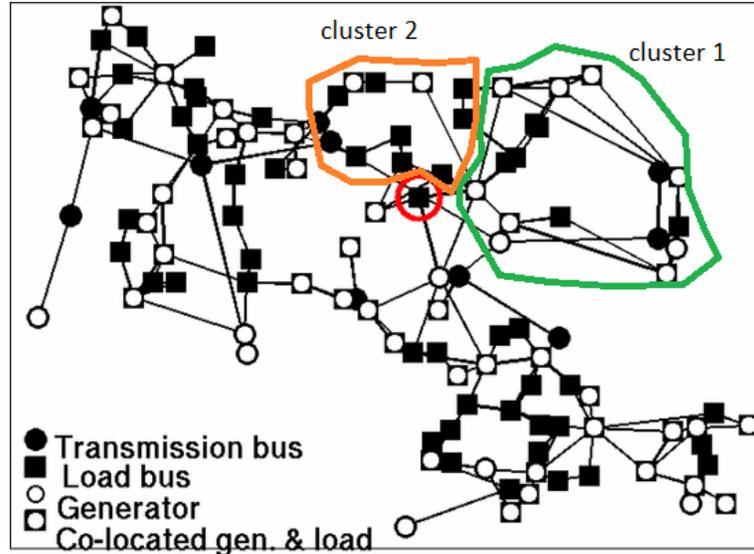


Figure 2.1: IEEE 118-bus topology (control-center marked). Green and orange mark defines the clusters we used for simulation in Section 2.4.3 [6].

grid infrastructure is that interdependencies among reliable systems, i.e., systems with exponentially distributed failure sizes, can make the overall system behave unreliably, as evidenced by power-law distributions for the overall system while individual layers' exhibited exponentially distributed failure sizes. To the best of our knowledge, a combined interaction among power, communication, and human layers' were not considered in any of the previous works relating to the cascading-failure studies of the power grid. As the critical contribution of this work, we analyze the coupled interactions among these three-layers' and develop a comprehensive model that captures the complex dynamics of the cascading-failures in the power grid.

2.1 Simulating cascading failures in power grids

The basic structure of the cascading failure simulation framework was developed in a prior dissertation work [80]. Due to the dependence of the simulation framework on the successive works, we first summarize the development of the simulation framework. The following link contains the simulation codes used in this

dissertation (<https://github.com/rashuvro/Modeling-Cascading-failures-in-smart-grids-Ph.D.-research->)

2.2 Cascading failure simulation (CFS) framework

We use MATPOWER [79], a package of MATLAB m-files for solving the steady-state DC/AC power flow optimization problem [81]. It uses the power-flow distribution framework under the given set of constraints. The standard power flow or load flow problem involves solving for the set of voltages and flows in a network corresponding to a specified pattern of load and generation [81]. MATPOWER includes solvers for both AC and DC optimal power flow problems, both of which involve solving a set of equations of the form $g(x) = 0$. In this dissertation, we develop our data set using the DC optimal power flow for simplicity. Here the word optimal refers that the power flow solutions depend on a set of constraints, and each time the best possible solution is provided satisfying the set of constraints. Although the AC power flow captures detailed dynamics of cascading failures in power grids, including the transient effects, the effect on the number of transmission line failures is found to be incremental [33].

2.2.1 Overview of the CFS framework

We show a flowchart in Figure 2.2 that illustrates the CFS framework used in this dissertation. We start with some initial number of transmission-line failure in the power grid initiated from an arbitrary initial event. The initial failures can be the outcome of a variety of initial incidents ranging from natural disasters to intentional sabotages. In subsequent work in this dissertation, we will show an approach for modeling these initial disturbance events. Here, it is important to note that we fail at least two transmission lines initially because the power grid is robust against one transmission line failure due to N-1 security considerations. Various incidents can trigger transmission line failures initially in the power grid,

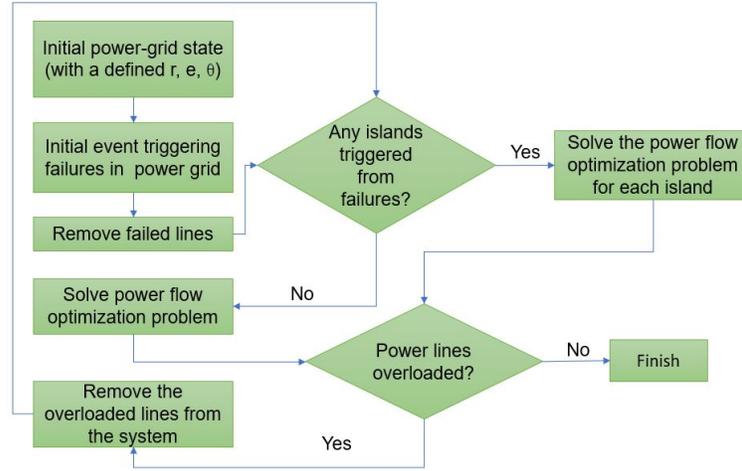


Figure 2.2: Flowchart of the cascading failure process in power grid

and we start our simulation using those failed transmission lines. The initial transmission line failures are randomly distributed over the space of the grid. We assume to have sufficient knowledge regarding the power grid topology and the operating parameters before the initial triggering event. We then remove the failed transmission lines from the system and check whether any islands are formed in the power grid because of those line failures. An island is a self-sufficient local network that operates independently when disconnected from the base network, having a set of generators and loads. Depending on whether any islands are formed or not, we then solve the DC optimal power flow using MATPOWER on each island. Note that, in the simulations, if the grids get islanded, we keep track all the islands (as if they are now individual grids) and run power-flow simulation in individual islands separately. Then we estimate the probability cascade stop and its parameters parametrically using the combined data obtained from all the islanded grids, not from the largest component of the grid. This approach implicitly captures islanding and its effects in the analytical model. For example, if a set of initial conditions result in various range of islands in the simulation, the estimated cascade stop probability would be different. If we have overloaded lines in the system, we can fail those lines, probabilistically fail a set

of lines among them (e.g., failing top N lines ranked using overload) or fail one transmission line per iteration of the power flow. In this dissertation, we use a similar approach used in [5] for islanding and overloading calculations and share two algorithms for calculating islanding and overloading in power grids. We repeat the same process until we end up with a system with no overloaded lines, which indicates the cascade-end.

This simulation framework is then used to perform Monte-Carlo simulations [82], which is a renowned computational simulation technique to obtain numerical results using random sampling. In Monte Carlo simulations, a subset of samples is chosen randomly from the population and then used to calculate the statistics of the population such as mean, distribution of the population. In our work, we use Monte Carlo technique to simulate cascading failures for various initial conditions of the grid chosen randomly, observe the cascading failure behaviors using the simulation framework described above, and then use the sample data to estimate the probability of cascade stop, which is a key parameter that governs the state transition probabilities of the Markov chain.

2.2.2 Power grid operating parameters and variables for cascading failures

Based on power grid simulations and prior works, we identify the following power grid operating parameters that govern the cascading failure dynamics. In our simulation, we use the IEEE 118-bus system (which is a simple approximation of the American Electric Power system (in the U.S. Midwest) [83]) as the test case which contains 186 transmission lines, 118 buses (nodes) and 54 generators.

power grid loading level, r : We define the power grid loading level, $r \in [0, 1]$ as the ratio of the total load demand, and the generation capacity of the power grid. In the IEEE 118-bus system, the maximum generation is 9966MW. $r = 1$ indicates the demand is 9966MW and $r \in [0, 1]$ scales the power demand with respect

to maximum possible generation. In our simulation, we simulate the cascading failures with various r . Note that a higher value of r increases the stress in the grid. We observe that for $r < 0.5$, the power grid is under no stress and can absorb the impact of two transmission line failures and redistribute the power flow without any further failures.

load-shedding constraint, θ : The load-shedding constraint is defined as the ratio of uncontrollable loads (loads that do not participate in load shedding) and the total load in the power grid denoted by, $\theta \in [0, 1]$. This is an important parameter to ensure the control actions by the power grid operator. $\theta = 1$ indicates that all the loads are uncontrollable, and the human operators can perform no load shedding. Again, $\theta = 0$ indicates that the operators can shed any load on the grid. In this paper, we consider equal load shedding constraints over all the loads in the grid. Further, we choose θ randomly between $[0,1]$ in our simulations. Similar to r , a higher value of θ increases the probability of cascading failure in the power grid.

Capacity estimation error, e : The Capacity estimation error, $e \in [0, 1]$ is defined as the error by the control center in its estimation of the actual capacity of the lines. In our CFS framework, this parameter is used to calculate overloaded lines. We used the same approach used in [5] to calculate overloaded lines. When power flow in a transmission line exceeds $(1-e) \times$ capacity, we consider that line as an overloaded line. We estimate the capacity of a transmission line using power flow simulation under maximum loads, i.e., when generation equals demand ($r=1$). Note that, since we use DC power flow simulation, there are no transient effects, and we can use the maximum generation without any issues. We quantize the flow capacity of a transmission line into a set of five capacities $\{20, 80, 200, 500, 800\}$ MW [81], and assign this capacity of the transmission line as a constraint of the MATPOWER power flow optimization problem (discussed later). In this paper,

we collect cascading failure data using various values of e .

In addition to the parameters mentioned above, we use the optimal power flow algorithm from MATPOWER, which includes the capability of implementing load shedding depending on cost. In our simulations, we set the cost of load shedding ten times higher than the cost of generation to ensure maximum generation before any load shedding. We track the cumulative amount of load shedding as a critical grid parameter. We keep track of the number of failed transmission lines and the maximum capacity of the failed lines during the propagation of cascading failures.

2.2.3 Optimal power flow

We solve the following DC optimal power flow equation 2.1 where F is a vector of power flow in transmission lines, A is a matrix whose elements can be calculated in terms of the connectivity of transmission lines in the power grid, and the impedance of the lines and P is a vector which contains the generator information.

$$F = AP \quad (2.1)$$

The optimization cost function and constraints are as follows,

$$\text{cost} = \sum_{i \in G} w_i^g g_i + \sum_{j \in L} w_j^l l_j \quad (2.2)$$

with the following optimization constraints

- (1) Power flow equations (1).
- (2) Generator power: $0 \leq g_i \leq G_i^{\max}, i \in G$
- (3) Controllable loads: $(1-\theta_j)L_j \leq b_j \leq 0$,
 $i \in L, l_j = b_j + \theta_j L_j$
- (4) Transmission line power flow: $F_k \leq C_k^{\text{opt}}$
- (5) Power balance: $\sum_{i \in G} g_i + \sum_{j \in L} l_j = 0$,

where g_i is the output of each generator, and l_i is the output at each load. C_k^{opt} is the capacity of a transmission line, and G_i^{max} is the generator capacity. In this cost function, w_i^g and w_j^l are positive values representing the generation cost and the load-shedding cost for every node, respectively. Controllable loads are defined using θ parameter. Finally, the power balance is done by the optimizer using a reference generator, typically the generator with the largest generation capacity is selected as the reference generator. The output of the MATPOWER contains power flow through each transmission line satisfying the constraints. If the load is greater than generation, then the excessive load is curtailed as load shedding. At this point, one might ask if the MATPOWER gives a power flow solution based on the set of constraints, then why there will be overloading? The answer is in the original MATPOWER solution; there is no overloading. However, to simulate cascading failure in power grids, we introduce a power flow estimation error $\in [0, 1]$ (discussed in detail later in this chapter), which controls the cascading mechanism. A higher power flow estimation error increases the chance of overloaded lines in the system and hence the probability of cascading failures. Again, the optimal power flow utilizes dispatchable loads to implement load shedding when the cost of generation is higher than the cost of serving loads. Here, since the load shed cost is set higher than the generation cost, load shedding is only performed when the optimizer fails to satisfy the other optimization constraints since analyzing the optimal generation and loads are not in the scope of this work.

2.2.4 Algorithms for finding the island and overloaded lines

Recall that we use DC optimal power-flow from MATPOWER, which is used in many cascading failure analysis papers [32, 29, 7, 5] for simplicity. However, we use a small failure probability (≤ 0.05) of the neighboring lines, which makes our model from deterministic to stochastic. We use the following algorithm for finding the maximum overloaded lines shown in algorithm 1.

Algorithm 1 Algorithm for finding maximum overloaded lines probabilistically.

Require: $\alpha, PF, Capacity$

Ensure: $FailedIndex$

```

1: for  $i \leftarrow 1$  to  $M$  do
2:    $P_{lf}(i) \leftarrow abs(PF(i))/((1 - \alpha) * Capacity(i))$ 
3:  $ProbTest \leftarrow 0$ 
4: for  $i \leftarrow 1$  to  $M$  do
5:   if  $rand < LinkProb(i)$  then
6:     if  $(ProbTest < P_{lf}(i))$  then
7:        $ProbTest = P_{lf}(i)$ ;
8:        $FailedIndex = i$ 
   return  $FailedIndex$ 

```

From the power flow data, overloaded transmission lines can be calculated, which is used in several previous works [84, 5, 1, 32] to fail transmission lines in the power grid. We consider a line failure when power flow through a line exceeds the maximum allowable power flow limit through that transmission line. Once we find overflow in a transmission line, we fail that line and re-calculate optimal power flow (OPF) using the remaining transmission lines. We take one or multiple transmission line failures per time unit to understand the cascading failure dynamics effectively. One way to choose one transmission line to fail out of all the overloaded lines is that if multiple transmission lines exceed the capacity threshold, we fail the line with the maximum deviation from the overflow threshold. Since the power grid needs to balance generation and load, the overloaded failed transmission lines can initiate a cascade of failures in the successive time steps.

Algorithm 2 shows our methodology for solving power flow in each of the islands created during the simulation. Recall that, an island in a power grid is a self-sufficient grid network containing both load and generators created when transmission line failure breaks the vast connected network into a small localized connected network. During each iteration, we calculate the number of islands formed in the grid and solve power flow simulation at each island using algorithm 2.

Algorithm 2 Algorithm for solving power flow in each islanded grid

Require: mpc, S, C

Ensure: GD, PF

```

 $SG = \text{sparse}(\text{AdjMatrix})$ 
 $[S, c] = \text{graphconncomp}(G)$ 
for  $i \leftarrow 1$  to  $s$  do
     $\text{struct } mpc[s] \leftarrow mpc$ 
     $[PF, GD] = \text{rundcopf}(mpc[s])$ 
return  $GD, PF$ 

```

2.3 Influence of communication network on the reliability of the power Grid

In this section, we study the influence of communication network functionality on the power grid. Specifically, we analyze the impact of power-communication interdependency on the reliability of the power grid during cascading failures. Two topological parameters of the communication network are studied, namely the hop distance and the node degree, which are used to determine the coupling between the two systems and to quantify the influence of the communication network on the power grid. We model the power-communication coupling parameter d , which captures the topological changes in the communication network. We study the influence of the power grid by analyzing the dependency between communication-link failures and transmission-line failures. The characterization of the coupling parameter, d is based on two topological parameters of the communication network, namely the hop distance and the node degree. The IDMC model in [44] refers to this coupling parameter but considered it as a constant. Instead, we formulate the coupling parameter d as a linear combination of the minimum hop distance and the maximum node degree of the failed communication nodes. Our observations illustrate that a decrease in the minimum hop distance or an increase in the maximum node degree of the failed communication nodes increases the probability of cascading failures in the power grid. By characterizing the coupling parameter, d , we analyze the impact of failures in the communication network on cascading-

failures in the power grid. Finally, numerical results have been carried out to quantify the impact of communication network failure on power grid reliability and validate our proposed work.

We start by adopting the IEEE 118-bus topology used in [6] for the transmission network of the power grid and use it to extract a topology for the communication network. In particular, communication nodes are placed at each substation, and communication links are set along each transmission-line similar to Supervisory control and data acquisition (SCADA) system, which is extensively used as a communication system in the power grid. Furthermore, there is one or multiple communication node(s) serving as a control center(s) in the power grid. Similar topologies of the power grid and communication networks were also considered in [7]. The control center monitors the health of the power grid and takes supervisory actions based on events to maintain optimal power flow. Figure 2.1 shows the control center (red circle) we used in IEEE 118-bus system. In this work, we have chosen the node with the largest node degree as the control center. Under normal operation, when disturbances occur in the power grid, operators at the control center can reconfigure the system by re-calculating the power flow and shedding a certain amount of loads. The load-shedding is usually done either manually or through an intelligent load-shedding management system over the communication network to balance the power flow [77].

We examine the dynamics of the communication network topology and its impact on the cascading-failure by including three key attributes describing the status of the communication network. The three attributes are the number of communication-node failures (y_n), the maximum degree of the failed communication-nodes (r_n), the minimum hop distance of the failed communication-nodes from the control center (h_n). Specifically, using [44], the state transition probability $f(s_{n+1}|s_n)$ from state $S_n = (x_n, i_n, y_n, l_n, k_n, h_n, r_n)$ to the next state $S_{n+1} =$

$(x_{n+1}, i_{n+1}, y_{n+1}, l_{n+1}, k_{n+1}, h_{n+1}, r_{n+1})$ is expressed below:

$$f(s_{n+1}|s_n) = \begin{cases} 1 & \text{if } i_n = 1, x_{n+1} = x_n, l_{n+1} = l_n, y_{n+1} = y_n \\ q(y_n) & \text{if } i_n = i_{n+1} = 0, l_n = 0, x_{n+1} = x_n, y_{n+1} = y_n + 1 \\ 1 - q(y_n) & \text{if } i_n = i_{n+1} = 0, l_n = 0, x_{n+1} = x_n, y_{n+1} = y_n \\ 1 - \frac{p(x_n)(1-d(y_n, h_n, r_n))}{(k_n + (1-d(y_n, h_n, r_n))(1-k_n))} & \\ & \text{if } i_n = i_{n+1} = 0, l_n = 1, x_{n+1} = x_n + 1, y_{n+1} = y_n \\ \frac{p(x_n)(1-d(y_n, h_n, r_n))}{(k_n + (1-d(y_n, h_n, r_n))(1-k_n))} & \\ & \text{if } i_n = 0, i_{n+1} = 1, l_n = 1, x_{n+1} = x_n, y_{n+1} = y_n \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

To this end, we express the communication network state at time n as a function of y_n , r_n and h_n . The newly introduced variables h_n and r_n further characterize the transition probability of the states, which were not considered in [44]. Note that, $q(y_n)$ is the probability of having one additional failure in the communication network.

Here, the quantity $d(y_n, r_n, h_n)$ indicates the coupling effect from communication network to power grid, and $p(x_n)$ is the probability of transiting to a stable state from a state with x_n transmission-line failures. Note that, $f(s_{n+1}|s_n)$ represents state transitions under different combinations of the state variables (includes newly added h_n and r_n) which are consistent with the IDMC model [44]. In the following section, we model the interdependency function $d(y_n, r_n, h_n)$, which captures the effects of the communication topology on cascading-failures. This is a major departure from the model reported in [44], where d was assumed to be a constant and the role of communication topology was not considered.

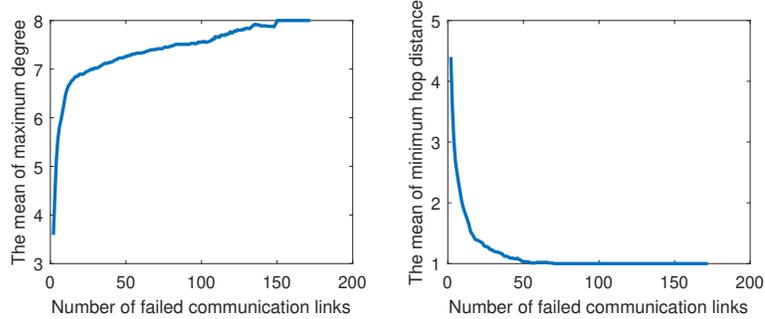


Figure 2.3: Empirical probability distribution, over the number of failed communication links, of the minimum hop distance of the center node to failed links (left) and the maximum degree (right) of failed nodes.

2.4 Modeling the communication-power interdependency function

In this section, we model the interdependency function $d(y_n, r_n, h_n)$.

2.4.1 Simulation setup

To evaluate the influence of the communication network on the power grid, we have developed a coupled communication and power-grid simulator. We conducted extensive Monte-Carlo simulations using MATPOWER [79], a package of MATLAB m-files for solving the optimal power flow problems. To simulate cascading-failures in power grids, we have used the approach described in chapter 2.1.

Specifically, the loading ratio, r of the power grid in the simulations, is assumed 0.7. The line tripping threshold, e , and the load-shedding constraint level, θ are assumed 0.3 and 0.2, respectively. Note that such a power-system operating setting makes the power grid under stress [5]. Now we initially fail a few random transmission-line failures in the power grid. Next, with probability $q(y_n)$ we simulate the associated communication-link failures which is adopted from the IDMC [44] model that includes a general term similar to $q(y_n)$. The probability $q(y_n)$ represents the influence of the power grid on the communication network. Following the simulation of the communication link failure, power flow through the

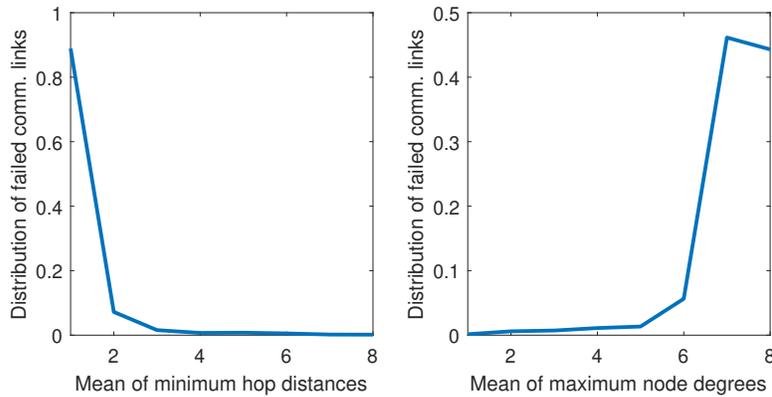


Figure 2.4: Empirical probability distribution of the communication-link failure over the minimum hop distance of the failed links (left) and the maximum node degree of the failed nodes (right).

transmission-lines in the power-grid is re-calculated. The overloaded line is failed using the line-failure algorithm described in chapter 2.1. In the simulation, we also assume that failures of communication links that are directly connected to the control center influence the failure in the power grid. Because, in a physical grid, the control center cannot send the necessary control signals to optimize the power flow through the missing communication path [7]. Communication links that are directly connected to the control center have a higher impact on reliability compared to other communication links as control centers are expected to monitor and interact with the electric devices remotely in real-time [74]. Hence, in the simulations, we scale up the probability of transmission-line failures in the power system based on the number of failed communication links that are directly connected to the control center. In the next subsection, the results of this simulation will be used to extract an analytical formula for the interdependency function $d(y_n, r_n, h_n)$.

2.4.2 Modeling the interdependency as a function of hop distance and node degree

Simulation results suggest that the communication nodes with higher degrees or lower hop distances have a greater impact on cascading-failures in the power-grid than the communication nodes with lower degrees or higher hop distances. In particular, the probability for the occurrence of cascading-failures in power grids increases when some communication node with high degrees or small hop distances from the control center fails. Here we develop a simple form, in conjunction with certain reasonable assumptions (originated from the results of the optimal power-flow simulation), to approximately represent the interdependency function, $d(y_n, r_n, h_n)$, whose range is in the interval $[0,1]$. The proposed form of $d(y_n, r_n, h_n)$ is:

$$d(y_n, r_n, h_n) = wp_{hop}^{fail}(h_n) + (1 - w)p_{degree}^{fail}(r_n) \quad (2.4)$$

where w is a weight factor between 0 and 1. In equation (2.4), the terms $p_{hop}^{fail}(h_n)$ and $p_{degree}^{fail}(r_n)$ refer to the probability distribution, which indicates the probability of a communication-link failure attributed to the minimum hop-distance state and maximum node-degree state, respectively. These probabilities can be estimated from the simulation study. Figure 2.3 shows the dependence of the minimum hop distance (left) and maximum node degree (right) with the number of failed communication links. We then compute the empirical probability distribution of communication-link failures as a function of the minimum hop-distance (between the central communication node and the failed communication links) and the maximum node-degree (of the failed communication nodes), as shown in Figure 2.4. Note that, in the right Figure of 2.4, when the mean of maximum node degree was eight, the probability value did not follow the increasing trend because of lack of enough samples. The observations in Figures 2.3 and 2.4 im-

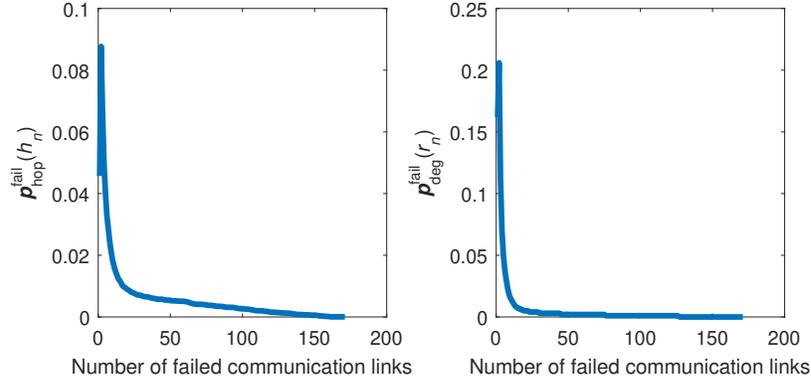


Figure 2.5: Hop distance (left) and node degree (right) distribution as a function of failed communication links.

ply that the interdependency function d on the state variables r_n and h_n can be made implicit through the variable y_n , i.e., $d(y_n)$. Hence, we can express $p_{hop}^{fail}(h_n)$ and $p_{degree}^{fail}(r_n)$ as $\tilde{p}_{hop}^{fail}(y_n)$ and $\tilde{p}_{degree}^{fail}(y_n)$, respectively. In order to completely characterize $d(y_n)$, we further need to have analytical formulas for $\tilde{p}_{hop}^{fail}(y_n)$ and $\tilde{p}_{degree}^{fail}(y_n)$. From Figure 2.4, we observe that the probability of communication-link failures, attributed to a certain minimum hop-distance (left) and maximum node-degree (right), respectively, follows a certain trend. Specifically, we can divide the observed behavior in each probability graph into two phases. For the minimum hop-distance probability (from Figure 2.4 (left)), we observed that at phase one, the probability of communication-link failure is decreasing from the maximum. While during phase two, the probability of communication-link failure is at minimum. Hence, we propose the following formula for $\tilde{p}_{hop}^{fail}(y_n)$:

$$\tilde{p}_{hop}^{fail}(y_n) = p_{hop}^{fail}(h_n) = \begin{cases} \frac{a_1}{h_n^4} & \text{if } 1 \leq h_n \leq m \\ \epsilon & \text{if } h_n > m \end{cases} \quad (2.5)$$

Here, we introduced the variable a_1 to represent the disturbance incurred when the minimum hop-distance of the communication system at unity. A minimum hop-distance value, $h_n = m$, indicates a critical value where the probability of

failure $\tilde{p}_{hop}^{fail}(y_n)$ decreases from a_1 to ϵ and enters from phase one to phase two (our results suggests $\epsilon = 0.01$).

From Figure 2.4 (right), the probability of failure at maximum node degree during phase one is minimal and equals to ϵ . A maximum node degree value, $r_n = n$, indicates a critical value where the probability of failure $\tilde{p}_{degree}^{fail}(y_n)$ exceeds ϵ (enters from phase one to phase two) and for values $r_n > n$, $\tilde{p}_{degree}^{fail}(y_n)$ increases monotonically. Hence, we propose the following formula for $\tilde{p}_{degree}^{fail}(y_n)$:

$$\tilde{p}_{degree}^{fail}(y_n) = p_{degree}^{fail}(r_n) = \begin{cases} \epsilon & \text{if } 1 \leq r_n < n \\ a_2 r_n^4 + \epsilon & \text{if } r_n \geq n \end{cases} \quad (2.6)$$

Similar to variable a_1 considered for the minimum hop-distance, we introduce a variable a_2 to represent the disturbance incurred from maximum node-degree of the communication system at unity. With equation (2.5 and (2.6) at hand, we have completely modeled $d(y_n)$.

To this end, the interdependency function d can be approximated by a function that depends on y_n alone. Generally, we note that $d(y_n)$ has the maximum value when the failed communication links have the minimum hop distance h_n and maximum node degree r_n for any given communication network. Intuitively, the lower the value of $d(y_n)$, the more stable the power grid is. Specifically, $d(y_n) = 0$ indicates that there is no dependency of communication-network failure on power-system failures. In contrast, $d(y_n) = 1$ implies deterministic failure in the power system, i.e., every failure of the transmission-line will cause a failure in the associated communication link. Throughout this paper we will assume $w = 0.5$, i.e., equal weights for simplicity.

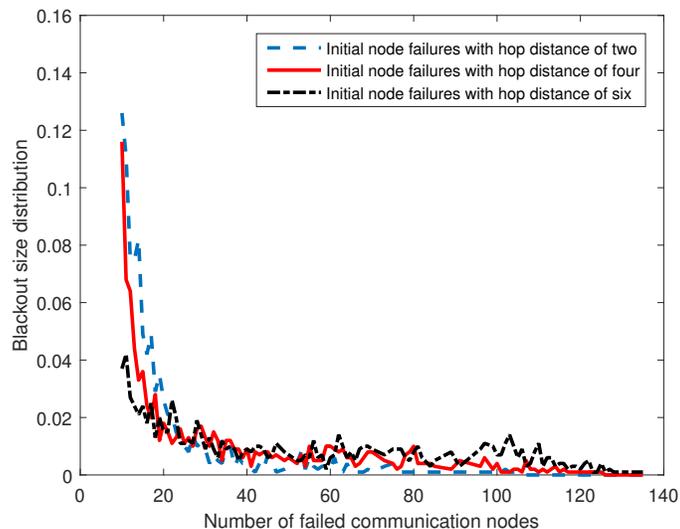


Figure 2.6: Blackout size distribution, ten initial random failures each with certain hop distance of two, four and six.

2.4.3 Simulation of the IEEE 118-bus system including communication network influence for different case

Using the aforementioned simulator, we have simulated the IEEE 118-bus system for different types of failure scenarios and empirically calculated the blackout distributions for each case. We have simulated the communication topology based on the following two particular scenarios.

First, we have simulated the blackout distribution in the power grid for ten initial communication node and power line failures with various hop distances. From Figure 2.6, we observed that failure in communication nodes with lower hop distances has higher blackout distribution as failures increase. This observation signifies that we have a higher probability of cascading-failure when failures occur at lower hop distances while keeping the node degree of the failed nodes fixed. Thus, we conclude that the blackout probability increases as the mean of the minimum hop distance are lower.

Next, we study the cluster failure scenarios. In particular, we will compare the

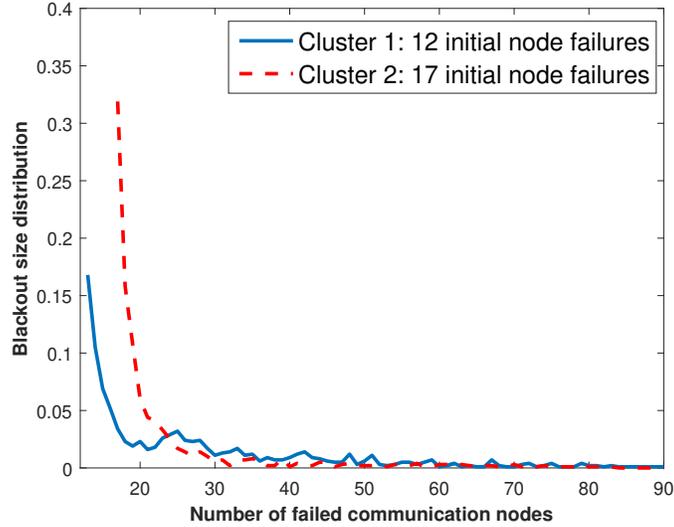


Figure 2.7: Blackout size distribution for the two cluster-failure scenarios.

blackout distribution for two clusters (clusters one and two are marked as green and orange, respectively), as shown in Figure 2.1. Note that both the mean of the minimum hop and the mean of the maximum degree are higher in cluster two to those in cluster one. As expected, we can observe in Figure 2.7 that cluster two is more conducive to cascading-failure than cluster one with higher blackout distribution during the initial phase of the communication-network node failures.

2.4.4 Simulating the Markov chain of the proposed model

With the proposed interdependency function d , as shown in equation (2.4), in conjunction with equation (2.5) and (2.6), we simulate the Markov chain of the proposed model. Our purpose is to validate the IDMC model by comparing its results to those obtained from the coupled communication and power-grid simulator. To do so, we study the case with three random initial transmission-line failures. Note that the topological property of the communication network is embedded in the interdependency function, d since d was extracted from simulation results using the actual communication network topology. Besides, the topological property of the power system is embedded in the transition stop probability $p(x_n)$

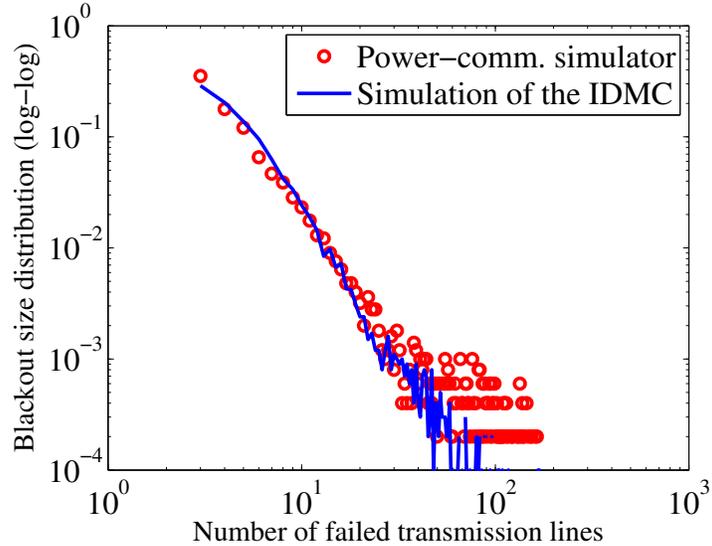


Figure 2.8: Blackout distribution (log-log scale) obtained from simulating the proposed model and the coupled power-to-communication simulator.

(established in the IDMC model [44] as a bowl shape function). In the simulation of the IDMC model, we simulate the evolution of the two coupled Markov chains based on the transition matrix $f(s_{n+1}|s_n)$, in conjunction with the interdependency function d proposed in Section V, for 10^6 iterations. For each iteration, the evolution stops when it reaches an absorbing state. We then average the blackout sizes obtained overall iterations to calculate the blackout size distribution. In Figure 2.8, we compare the blackout distribution of cascading-failures obtained empirically by simulating the proposed model with the results obtained from the coupled communication and power-grid simulator. It is clear that the two results agree in showing a similar trend in the blackout size distribution. Note that the results obtained from the coupled simulator are not precise when the number of failed transmission-lines is large (e.g., over 100), which is due to the limited sample size of large blackouts. All in all, these results validate that the proposed model is effective in capturing the impact of the interdependency between the power system and communication network on cascading-failures in the power-grid. Thus, we have developed a communication-power interdependency function, d that is

determined by hop distance from a central node, and the degree of the node in the communication network. This coupling function captures the influence of the communication networks on the power system under different stress levels of the power grid during cascading-failures. We have devised a coupled power-communication simulator and conducted extensive simulations to validate the proposed modeling of the coupling function in various cascading-failure scenarios. A key insight obtained from the simulation results is that the total blackout probability in the power grid can be significantly impacted by the failures in the communication network when the power grid is under stress. Our simulation results illustrate that the proposed model is an efficient model to investigate the cascading-failures in the interdependent power grid. The computational time for simulating the proposed model is reduced by a factor of 10^7 to the time using the coupled simulator.

2.5 A Markov-chain based three-layer Model for cascading-failures in smart grids

In this section, we refine the SASE model to capture the influence of the communication network and human operators' response on the reliability of the power grid. The Markov chain captures the cascading-failure dynamics in the power grid in reduced abstract state-space. Each state in the Markov chain corresponds to a specific state of the power grid. We define the states at which cascading-failure ends as the absorbing states. When the Markov chain reaches to an absorbing state, there will be no further failures. The rest of the states of the Markov chain is termed as cascade-continue states in which cascading-failure will evolve until it reaches an absorbing state. We also characterize the state transitions for the Markov chain model in this section using our simulations. Finally, we study the blackout sizes of the power grid given any initial condition.

2.5.1 Markov chain based cascading-failure model capturing power, communication and human factors

Recall that, in the Stochastic Abstract space evolution (SASE) model [5], cascading-failure evolution was characterized using a Markov chain, which captures the failure dynamics in the power system. In particular, in the SASE model, a state S_i of the Markov chain consisting of three state variables: the transmission line failures in the system, X_i , maximum capacity of the failed transmission lines C_i^{max} and a $\{0, 1\}$ -valued variable termed cascade-stability of the system, I_i , for which $I_i = 0$ indicates the system is susceptible to further failures and $I_i = 1$ indicates that no further failures are possible. Hence, the system has entered a cascade-stable mode. The operating characteristics considered in the SASE model [5] are based on power grid simulations, and they include line-tripping threshold, e , power-grid loading level, r , and the load-shedding constraint level, θ . Power grid is stressed when the characteristic parameters, r , e , θ are high. Cascade-stop probability $P_{stop}(S_i)$, was characterized parametrically in the SASE model using power system simulations [5]. In work by Wang *et al.*, a new state variable for the Markov chain, i.e., human operators' response H_i , was introduced [85]. A coupled parameter $g(H_i)$ was introduced to capture the influence of the human operator on the power grid. Cascade-stop probability $P_{stop}(S_i)$ was defined parametrically using an operating characteristic of the power grid while considering the role of human factors. In the previous section, the power-communication coupling parameter, d is modeled that influence of the communication network on the power grid. It has been shown that a decrement in the minimum hop distance or an increment in the maximum node degree of the failed communication nodes increases the cascading-failure probability in the power grid. To this end, we observe from [5, 44, 62, 85] that using a Markov chain the coupled interactions between different

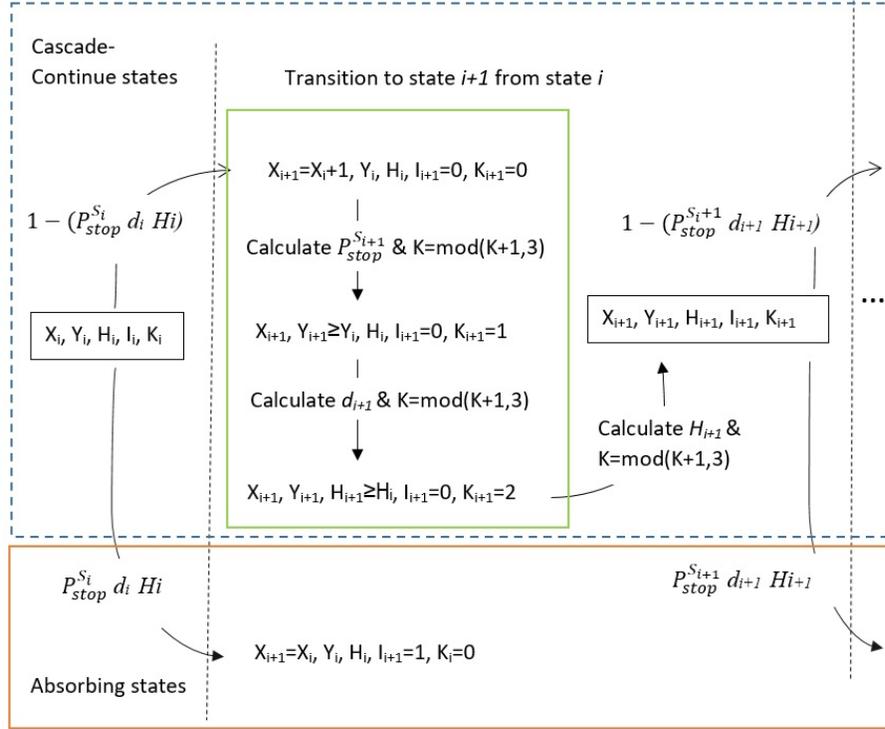


Figure 2.9: State transition diagram for the three layer Markov chain

layers' of the power grid can be explained efficiently.

In this section, we consider IEEE 118-bus topology, which contains 186 transmission lines. Similar to the previous section, we consider the same topology for the communication network, which has a one-one mapping between power and communication nodes. In particular, communication nodes are placed at each substation, and communication links are set along each transmission line. Furthermore, there is one or multiple communication nodes serving as a control center(s) in the power grid. The power grid operation control center monitors the health of the power grid and takes supervisory actions based on various events to maintain optimum power flow in the power transmission layer. For example, Figure 2.1 shows the IEEE 118-bus topology for the power grids and a marked control center (red circle) in the IEEE 118-bus system. During contingencies, when power grid disturbances occur, human operators in the control center can reconfigure the system by, for instance, reconfiguring power flow, changing generators' set points, and

Table 2.1: Human Operator response levels based on failure in power and communication network

Level	Definition	Available action time	Stress
1	$X \leq 5$ and $Y \leq 10$	Normal	Normal
2	$5 < X \leq 10$ and $10 < Y \leq 30$	Low	High
3	$10 < X \leq 50$ and $Y > 30$	Extremely low	Extreme
4	$X > 50$	N/A	N/A

shedding a certain amount of load. Load-shedding is usually done either manually or through an intelligent load-shedding management system (while the communication network plays a key role in implementing the load-shedding decisions) to balance the power flow [77]. In the three-layers of the Markov chain, we consider that the absorbing state is associated with the power grid layer because, once a cascading-failure is initiated, it can stop only when there are no further failures in the power grid. We evaluate the parameters ($P_{stop}(S_i)$, d_i and H_i) of the current state (i.e, i) to calculate the state transition probability (described in the next sub-section) of the next state of the Markov chain.

2.5.2 State transitions

We consider the state variables, $S_i = (X_i, Y_i, H_i, I_i, K_i)$ in the three-layer model to represent an abstract state S_i , where X_i is the number of transmission-line failures in power grid; Y_i is the number of communication-link failures; H_i refers human operator performance level; $I_i =$ power grid stability indicator, a $\{0,1\}$ valued binary variable, and K_i is the layer tracking parameter. K is defined as $K = \text{mod}(k, 3)$, $k = 0, 1, 2, \dots$ for the three-layers' and k increases one step at a time. At $K = 0$, the power grid dynamics are evaluated, at $K = 1$, the communication network dynamics are assessed, and at $K = 2$, we estimate the human operator performance level. We use the coupling parameter, d for communication networks influence as described in the previous section, and coupling parameter for human

operation's influence, $g(H_i)$ from [85].

Figure 2.9 illustrates the state transitions in the three-layers' characterizing cascading-failures in the smart grid. We did not consider any healing during cascade propagation, i.e., failures increase monotonically in the power grid (i.e., X_i is non-decreasing) and have focused on the negative effects of interdependencies in this work. We consider that when failures trigger in the power grid, failures in the communication network and the operator response level (leads to human operator error) also increases monotonically during cascading-failures. From Figure 2.9, we observe that, for a transition from S_i to S_{i+1} state, we go through each layer of the three-layer system using the layer tracking parameter, K (power grid, communication network, and human operator response) and evaluate the coupling parameters. We start the coupling parameters with some initial value conditional on the initial state (for example, if power grid and communication network is operating with no failures, then $d = 0$ and $g(H_i) = 0$). We calculate $P_{stop}(S_i)$ with X_i failures using the following analytic formulation from [5].

$$P_{stop}(S_i) = \begin{cases} a_1 \left(\frac{a_2 L - X_i}{a_2 L} \right) + \epsilon & \text{if } 1 \leq X_i \leq a_2 L \\ \epsilon & \text{if } 1 < X_i \leq 0.6L \\ Q(X_i) & \text{if } 0.6L < X_i \leq L, \end{cases} \quad (2.7)$$

where L is the total number of transmission-lines, $Q(X_i)$ is a fixed quadratic function approximating the tail of the family of bowl-shape functions. In our simulations for the three-layer model we also got similar bowl-shape behavior as illustrated in Figure 2.10. Without loss of generality $P_{stop}(L) = 1$, i.e., there is no more transmission-line failures when all the transmission lines have failed in the power grid. Scaling parameters a_1 , a_2 and ϵ were approximated using the power grid characteristic parameters (r, e, θ) using the following parametric formulation obtained from power system simulation presented in [1]:

$$a_1 = 0.4 - 0.25[[r]]_{0.5} - [e]_{0.1}^{0.5}(0.2 - [e]_{0.1}^{0.5}) - 0.25[[\theta]]^{0.4} \quad (2.8)$$

$$a_2 = 0.1 - 0.05[[r]]_{0.5} - 0.1[e]_{0.1}^{0.5}(0.2 - [e]_{0.1}^{0.5}) - 0.07[[\theta]]^{0.4} \quad (2.9)$$

$$\epsilon = 0.6 - 0.4[[r]]_{0.5} - 0.5[e]_{0.1}^{0.5} - 0.3[[\theta]]^{0.4}. \quad (2.10)$$

Note that, the constants in the scaling parameters a_1 , a_2 and ϵ are empirically calculated based on the simulation results (part of the results have been shown in [6, 5, 1]). When $X_{i+1} = X_i$; $I_i = 0$; $I_{i+1} = 1$; $K_{i+1} = 0$; the Markov chain will have a state transition from a cascade-continue state to an absorbing state with probability $P(S_{i+1}|S_i) = P_{stop}(S_i) d_i g(H_i)$ (marked with solid line in Figure 2.9). Here, we took the Cartesian product among the coupling parameters as all the three coupling parameters are $[0,1]$ valued probability estimates and Cartesian the product (which is also $[0,1]$ valued) gives a combined state transition from a cascade-continue state to an absorbing state. Note that, when there are no failures in the power grid, $d = 1$ and $g(H_i) = 1$.

For $X_{i+1} = X_i + 1$; $I_{i+1} = I_i = 0$; $K_{i+1} = 0$; the Markov chain will be in a transitory state (marked with dotted line in Figure 2.9) having one new failure in the power grid, with following transition probability: $P(S_{i+1}|S_i) = 1 - (P_{stop}(S_i) d_i g(H_i))$. At this stage, we will update the coupling parameters from each layer for the state $i + 1$. The status of the current layer is tracked using the layer tracking parameter, K . With new failures in the power grid and with $K = 0$, we calculate $P_{stop}(S_{i+1})$ using equation (2.7) and update the layer tracking parameter by one step using $K = mod(k + 1, 3)$. At this stage ($K = 1$), we will evaluate the failure dynamics of the communication network and calculate the communication-power coupling parameter (d) in terms of minimum hop distance of the failed communication lines and maximum degree of the failed communication nodes using equation 2.4. Now, at $K = 1$, if $Y_{i+1} = Y_i$; (i.e. there is no fail in the

communication network) then $d_{i+1} = d_i$ and if $Y_{i+1} > Y_i$; we calculate d_{i+1} using (2.4) and update the layer tracking parameter by one step using $K = \text{mod}(k+1, 3)$.

At this stage ($K = 2$), we will evaluate the dynamics of the human operation involved in the cascading-failure process and calculate the operator response level H_i using the hSASE model [1]. Table 5.1 shows the relationship between the human operator performance level with the power grid and communication network. We characterize the relation between human operator response due to failure in the power and communication network in four levels. Here, level 1 indicates normal power grid operation, and level 4 indicates certain failure dynamics when operators' cannot stop the cascading-failure due to an excessive number of failures that have happened already. But the authors in [1] did not consider the role of communication topology while calculating human operator response. Using the definition from Table 5.1 (which includes failure in power grid and communication network), we then calculate the HEP using the following formula developed in [64] (interested readers can refer to [63] [64] where human error probabilities and their interactions with performance shaping factors (PSF) and nominal HEPs (NHEP) were described in detail)

$$HEP(H_i) = NHEP \prod_{i=1}^2 PSF_i. \quad (2.11)$$

In our model, two PSFs have been considered (available action time and stress level of the operator) for calculating $HEP(H_i)$. When $H_{i+1} = H_i$, then there is no change in human operator performance level and If $H_{i+1} > H_i$; the human operator stress level is increased and we calculate the new performance level, H_{i+1} using $g(H_{i+1}) = 1 - b \cdot HEP(H_{i+1})$. Here, b is a free scaling parameter [1].

To this end, we have completely evaluated the three coupling parameters that influence each other. Now, based on whether there will be any new failures in

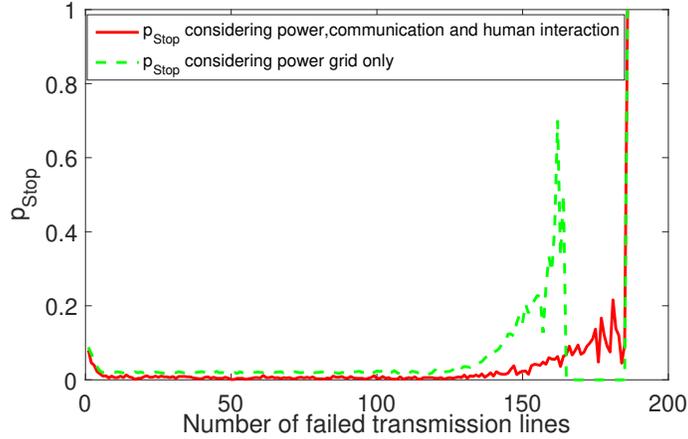


Figure 2.10: Cascade-stop probability with and without the influence of communication and human response.

the power grid or not, we again go to an absorbing or transitory state using the state transition probability and calculate the coupling parameter for the new state again. This loop of state transition and updating the coupling parameters continues until the Markov chain goes into an absorbing state.

2.6 Simulation Results

Using our parametric formulation of the state transition dynamics, we now perform Monte-Carlo simulation with the Markov chain. We consider the IEEE 118-bus power grid topology, which contains 186 transmission lines. We consider two failures in the power grid and communication network initially (to negate the N-1 security effect). We assume that if a communication node is failed, the links connected to that communication node will fail too. The stress level in the power grid is determined using the power grid characteristics parameter, r (power grid loading ratio), e (capacity estimator error), θ (load-shedding constraint), which was introduced in the SASE model [5]. We take blackout size distribution as the critical parameter to assess the model, which is a widely adopted metric for assessing cascading-failures [61, 31, 5]. It refers to the blackout probability of certain transmission lines after the Markov chain enters into an absorbing state

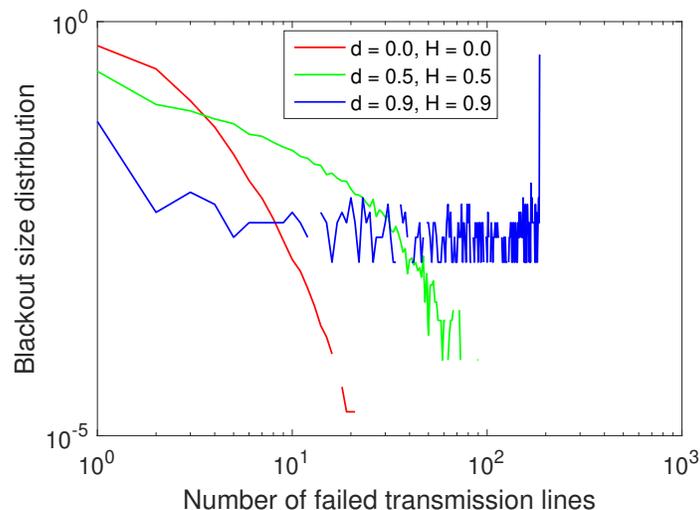


Figure 2.11: Blackout size (log-log scale) with and without the impact of communication and human operator influence.

under given initial condition (number of initial failures and power grid operating characteristic parameters, etc.).

First, Cascade-stop probabilities (with and without the consideration of the coupling from a communication network and human factors) as a function of the number of failed transmission lines are shown in Figure 2.10. At this state of the Markov chain, there will be no further failures, and cascading-failures in the power grid will stop. The cascade-stop probabilities were calculated when the power grid is under stress ($r = 0.95$, $e = 0.25$, $\theta = 0.25$). From Figure 2.10, it can be observed that cascade stop probability shows bowl shape for both the cases, which we discussed already in the previous section. However, when the communication network and human operator response were considered, we observed low cascade-stop probability as compared to the power-grid-only scenario for the transmission lines [5]. It illustrates poor human operator performance, and failure in the communication network can lead to greater blackout size.

Figure 2.11 represents the impact of blackout size for three cases (with no influence, moderate influence, and strong influence from a communication network and

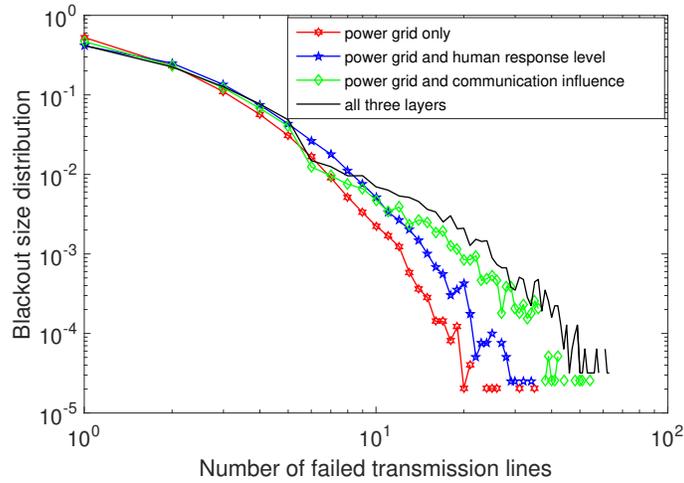


Figure 2.12: Blackout size (log-log scale) comparison when the grid is not in stress ($r = 0.5$, $e = 0.1$ and $\theta = 0.1$ was considered for no stress scenario).

human operator into the power grid). Here, we considered the coupling parameter d and H as 0.5 for moderate influence and 0.9 for high influence case. We observe that without any influence, the blackout size follows an exponential distribution, while for high deterministic influence, the blackout size follows a power-law distribution. The results obtained from the coupled simulator are not precise when the number of failed transmission lines is large (e.g., over 100), which is due to the limited sample size of massive blackouts.

Figures 2.12 and 2.13 represent the probability of blackout when the power grid is under no stress and under stress, respectively (stress is regulated using the characteristic parameters of the power grid). We consider four cases, i.e., power grid only, power and communication coupling, power and human operator response coupling, and all three-layers' coupling to analyze the failure propagation and for blackout size calculation. We observe that blackout size increases significantly when the power grid is stressed and influenced by the communication network and human operator response. It is visible from our simulation results that the blackout size becomes more substantial when the power grid is in stress, and there are influences from a communication network and human operators'. Figure 2.13

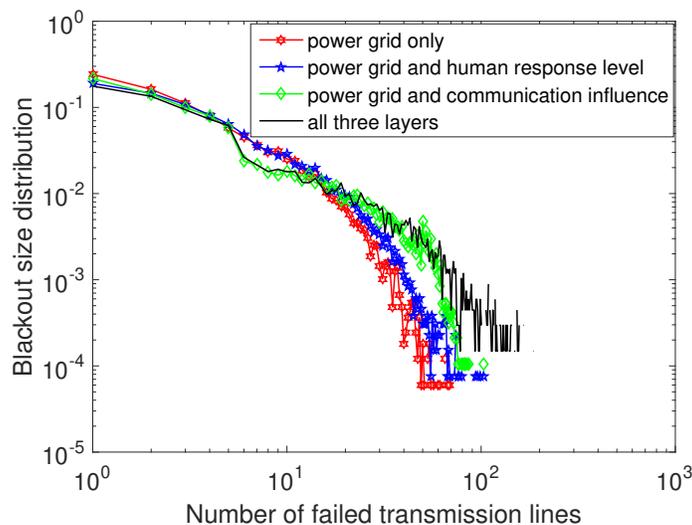


Figure 2.13: Blackout size (log-log scale) comparison when the grid is under stress. ($r = 0.8$, $e = 0.2$ and $\theta = 0.2$ was considered for stressed scenario).

suggests that blackout size for the three-layer follows power-law distribution after certain power grid operating parameters, although, for the single and two-layer interactions, it shows the exponential distribution. From Figure 2.13, we can see that blackout size distribution for all three-layer (black), and power grid and communication influence (green) starts leaning to power-law when the number of transmission line failure is around 50 while in the other two cases it's exponential. This is because of the higher influence of communication node failure than human error. So, blackout sizes can result in reliable (exponential) and unreliable (power-law) behavior due to different characteristic parameters' and coupling between the 3-layers.

In this chapter, we discussed the role of communication and human operator response level in the cascading failure dynamics in smart grids. However, the detailed state transition probabilities of the Markov chain are not captured in this work, which is a major drawback of the work. Again, as each layer of the grids is considered to have an individual Markov chain, as the network grows, scalability and tractability becomes a major concern with the IDMC approach, Which leads

us to develop an alternative solution to be discussed in the next two chapters.

Chapter 3

Correlating grid-operators' performance with grid variables in smart-grids during cascading failures

In this chapter, the role of human factors associated with the grid operators, e.g., human-error probability (HEP), is modeled as a function of the grid conditions as well as operators' training and experience levels. Moreover, the HEP is embedded in a previously reported Markov-chain model that generates the probability distribution of blackout as a function of time following a trigger. Specifically, through the HEP, the Markov-chain's transition matrix includes the dynamics of detailed smart-grid operator attributes. To derive the grid-state dependent HEP, three real-valued performance shaping factors (PSFs), representing key human attributes of the operators, are mapped to the grid-state variables, thereby capturing the correlation between the evolution of the PSF levels and the propagation of transmission-line failures. This mapping is established based on a histogram-equalization principle, which utilizes the experimentally-estimated probability distribution of the PSF levels while assuming a monotone relationship between the HEP values and number of line failures. Further, the distribution of the PSF levels was used to identify the critical combinations of PSF levels that correspond to an event with high joint probability as well as a high HEP.

3.1 Introduction

Human operators play a pivotal role in mitigating the propagation of cascading failures. Human operator error (HEP) during propagation of failures can increase the probability of a large blackout astronomically. Historical data analysis of large power-grid blackouts strongly indicates operator error as a critical initiator of

cascading failures. The 2003 Northeast blackout in the United States and Canada occurred due to a combination of transmission-line and generating-unit failures, communication components, and server failures, and ineffective and erroneous human-operator responses [4]. Hence, studying the interplay between the power grid and the smart-grid operators during cascading failures is critical in predicting the reliability of the smart-grid. [15].

In this chapter, we propose a Markov-chain based cascading failure model, including human operator actions and decisions in the loop. The joint probability of three PSFs (available time, stress, and complexity of the problem) is used to map the grid operators' response levels with the grid states of the Markov chain. The other five PSFs (experience of the operators, work process, fitness, ergonomics, and procedures), which are not affected by the propagation of failures, is used as a fixed initial parameter of the Markov chain for calculating the HEP, i.e., three PSFs will have varying PSF levels (depending on the grid state of the Markov chain) multipliers, and five PSFs will have a fixed PSF level and the associated multipliers during the propagation of failures. Then, the calculated HEP from the PSFs in a grid state is used to evaluate the probability of cascade stopping (in the Markov chain), including the role of human operators. Notably, this incorporation of the correlated and uncorrelated PSFs using our proposed methodology drastically changes the state transition probabilities of the Markov chain compared to [1], which now captures the detailed dynamics of the role of smart-grid operators during cascading failures in the power grid. Finally, we identify a set of critical PSF level combinations from all the possible events consisting of various combinations of PSF levels. We show that only considering the HEP without considering the distribution of the PSFs can be misleading, as in most cases, a combination of PSF events will lead to high HEP but with very small/zero probability of occurrence.

3.2 Markov chain based cascading failure model including operator in the loop

3.2.1 State variables and transition matrix of the Markov chain

Using a similar approach reported in [5, 1], the detailed power-grid states in the same class are represented by a few aggregate state variables. The state variables defining each equivalence class are the number of transmission-line failures, F_i , the maximum capacity of the failed lines C_i^{max} and a human-factor variable, H_i , which captures the status of operators' performance (using PSFs) controlling the power grid. Notably, the reduced state variables also include a critical variable, I_i , to capture the complex event of the cascade stopping: if the power grid is in a cascading mode, then $I_i = 0$ and the cascade will continue. Conversely, the cascading failure terminates if the power grid is in an absorbing state, namely when $I_i = 1$. The state-dependent cascading-stop probability, P_{stop} , is parametrically expressed in terms of power-grid loading level, r , capacity-estimation error, e , and load-shedding constraint, θ [5].

The transition matrix of the Markov chain is a $2M|\mathcal{C}| \times 2M|\mathcal{C}|$ matrix, where M is the total number of failed transmission lines and $|\mathcal{C}|$ is the cardinality of the set of capacities, which in this work is five, and 2 accounts for the binary variable I . In our model, we do not consider any healing capability, thus, transmission-line failure, F_i , increases monotonically with one failure per unit step. The state transition probability, $f(S_j|S_i)$, from state $S_i = (F_i, C_i^{max}, H_i, I_i)$ to the next state

$S_j = (F_j, C_j^{max}, H_j, I_j)$, is defined below

$$f(S_j|S_i) = \begin{cases} 1 & \text{if } F_j = F_i, C_j^{max} = C_i^{max}, I_j = I_i = 1, \\ P_{stop}(S_i)g(H_i) & \\ & \text{if } F_j = F_i, C_j^{max} = C_i^{max}, I_j = 1, \\ P(S_j|S_i) & \\ & \text{if } F_j = F_i + 1, C_j^{max} \in \mathcal{C}, I_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

Here $P_{stop}(S_i)$ is the probability of cascade-stop at state S_i , and the human factor variable is mapped explicitly using the F_i and C_i^{max} , which is described in the following section. We adopt the formulation of $P_{stop}(S_i)$ from [1] and $P(S_j|S_i)$ from [5] (for details see [5, 1]). Note that $g(H_i) = 1 - \text{HEP}(H_i)$ is a function that translates HEP for a specific human operator response into the transition matrix of the Markov chain. Next, we describe the formulation of HEP.

During the initial triggering phase (precursor phase) of the cascading failures in the power grid, the status of several power-grid operating parameters (i.e., loading level, the capability to implement load shedding, etc.) is critical to trigger cascading failures and to determine the size of the blackout. In addition to these power-grid operating parameters, the human operator attributes during the initiation of a cascade play a critical part. For example, in most of the contingency scenarios, operators are equipped with diagnostic procedures to devise a mitigation strategy, have sufficient time to analyze the problem, and implement actions accordingly. In [64], the authors used Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) methodology to calculate the HEP for a given context. The SPAR-H methodology is a simplified approach to human error quantification that accounts for individual performance-shaping factors that affect

perception, processing and response to events in complex environments [63]. Co-author Dr. Abreu interviewed system operators from New England and the South East, and took note of the narratives of the emergencies that happened during their shift, which is then applied to the SPAR-H methodology to each task and calculated dependency between the cognitive and action-based tasks. Then they calculated the final failure probability and compared the empirical evaluation of error probability with what actually occurred. Using Monte Carlo methods and the SPAR-H equations, final estimation of the HEP and the frequency associated with each level was calculated. Detailed step by step SPAR-H description of the process can be found in [86]. Here the HEP is formulated using the PSFs of the operators. The PSF multiplier values for a specific operator context was listed in Table 1 of [64]. Note that, during the cascade triggering phase, each operator dealing with the scenario can possess a specific set of attributes. The probability that an operator will take a good/bad decision depends on the specific set of attributes. These attributes refer to various PSF levels in Table 1 in [64]. Specifically, the human operator attributes were divided into eight different PSFs, each containing a set of PSF levels. Each PSF level is associated with a multiplier value, which is used to calculate the HEP quantitatively using (4.2) (adopted from [64]).

$$\text{HEP} = \frac{\text{NHEP} \prod_{i=1}^8 \text{PSF}_i}{\text{NHEP}(\prod_{i=1}^8 \text{PSF}_i - 1) + 1} \quad (3.2)$$

Here NHEP represents the type of operation (diagnosis or action, having a multiplier for each type) performed by the operators. Further, the distribution of the PSFs (calculated based on grid-operator interviews) are obtained from Table 3 in [64].

3.2.2 Mapping between PSFs and the Markov chain state-space

A key objective of this work is to find a data-driven mapping between power grid states and the PSFs using the distribution of the PSFs. The operator attributes can be characterized using various combinations of the PSF levels, and depending on the combinations, the HEP during the initial triggering phase of a cascading failure event can vary between zero and one. The PSFs such as available time, stress, and complexity are correlated with the grid conditions. Specifically, it is intuitive that as the failures propagate, available time to react for an operator would be less, the stress on operators to mitigate the contingency would be high, and the complexity of the problem would become more complicated. On the other hand, performance shaping factors such as ergonomics, the fitness of the operators, availability of the procedures, experience of the operators, and work processes do not change during the propagation of the failures. Although these factors do not correlate with the propagation of failures, they are critical and play a significant role in quantifying the HEP. For example, from Table 1 in [64], when the available time is inadequate, or the operator is unfit to work, the human error probability is one, which indicates that an unfit worker or an operator with inadequate time would certainly make an error.

As described in the related works, Wang *et al.* used two PSFs (available time and stress) and coarsely mapped grid states with the operator levels (Table 2 in [1]). However, as the operators' response is quantized in four operating levels only (each associated with different PSF multipliers as shown in Table 2 in [1]), there is a limited variation of the human error probability. To overcome this serious limitation, we use all the eight PSFs reported in [64] and their associated multipliers for calculating the HEP. Specifically, we judiciously use the three PSFs (available time, stress, and complexity) that are correlated with the propagation of

failures to define the operator response levels using the power-grid state variables. Note that available time, stress, and complexity has four, three, and four PSF levels, respectively. Thus, we have forty-eight ($4 \times 3 \times 4$) distinct combinations of these PSF levels, i.e., forty-eight distinct operators' response levels compared to only four responses in [1]. The distribution of each PSFs is available in [64]. The joint probability mass function of the PSF's is calculated using the following equation:

$$F(\text{PSF}_1, \dots, \text{PSF}_n) = P\{\text{PSF}_1 \leq \text{psf}_1, \dots, \text{PSF}_n \leq \text{psf}_n\} \quad (3.3)$$

Here $\text{psf}_1, \dots, \text{psf}_n$ represents the PSF levels for an individual PSF. In this work, we assume that the PSFs are uncorrelated, i.e., the likelihood that a PSF is independent of the other PSFs. Thus, we can calculate the joint probability mass function from their marginal mass functions using the following equation:

$$F(\text{PSF}_1, \dots, \text{PSF}_n) = \prod_{i=1}^n F(\text{PSF}_i) \quad (3.4)$$

Using (3) and (4), we calculate the joint probability for each of the forty-eight combinations of PSFs. To map these events into the state space of the Markov chain, we first determine the size of the Markov chain from the state variables. Similar to [5, 1], we use the IEEE 118-bus system, which has 186 transmission lines. The flow capacity in transmission lines is quantized to obtain a set of five transmission line capacities, $\mathcal{C} = \{20\text{MW}, 80\text{MW}, 200\text{MW}, 500\text{MW}, 800\text{MW}\}$, and each transmission line is assigned with capacity according to their power-flow capacities. Including the binary absorbing/continuity variable, the state space of the transition matrix is 1860×1860 where the index of a specific state of the Markov chain can be calculated as using, $(F_i - 1)|C_i^{max}| + 2(C_i^{max} - 1) + I_i + 1$.

To map the forty-eight operator response levels to 1860 grid states, we use the

Table 3.1: Mapping of Operator response levels with PSF levels and grid variables (nominal PSF levels considered for the other PSFs)

Operators response	available time	stress	complexity	HEP	joint probability	grid state(F_i, C_i^{max}), index
level 1	expansive time	nominal	obvious diagnosis	0.0000	0.0000528	(1,20), 1
level 2	expansive time	high	obvious diagnosis	0.0000	0.0000248	(1,20), 1
level 3	expansive time	extreme	obvious diagnosis	0.0001	0.000024	(1,20), 1
level 4	expansive time	nominal	nominal	0.0001	0.0000924	(1,20)-(1,80), (1-3)
level 5	extra time	nominal	obvious diagnosis	0.0001	0.001584	(1,200), 5
...
level 47	barely time	high	moderately complex	0.5025	0.008442	(185,500), 1847
level 48	barely time	extreme	highly complex	0.7163	0.001386	(185,800), 1849

Table 3.2: Detailed Grid variable and HEP mapping

	F=1	F=2	F=3	...	F=185	F=186
$C^{max} = 20$	0.0000	0.0010	.0010	...	0.5025	1
$C^{max} = 80$	0.0001	0.0010	0.00105025	1
$C^{max} = 200$	0.0002	0.0010	0.00105025	1
$C^{max} = 500$	0.0006	0.0010	0.00105025	1
$C^{max} = 800$	0.0010	0.0010	0.00107163	1

histogram equalization technique [78]. First, we calculate the HEPs for the operator response levels. Then we sort HEP from low to high. We assume that HEP increases monotonically with the grid index, i.e., more failures in the grid, and the higher value of maximum capacity will increase HEP. Then we multiply the joint PSF probability with 1850 (maximum StateIndex10 for IEEE 118-bus system) and round it to the nearest integer value to map the operator response levels to the grid StateIndexes as shown in Table 5.1. Finally, when all the transmission lines fail, i.e., for the last ten StateIndexes, the HEP of one is assigned since all the lines have failed already. Note that we only take the odd grid StateIndexes since the even states represent the absorbing states. Finally, if multiple operator responses are mapped to a single grid state (due to very low joint probability), we assign the average HEP (of those response levels) for that state. With this, the detailed grid state mapping with HEP is complete, and a snapshot (with nominal multipliers of the five uncorrelated PSFs) is shown in Table 3.2.

At the starting state, the grid operators would be attributed to various PSF

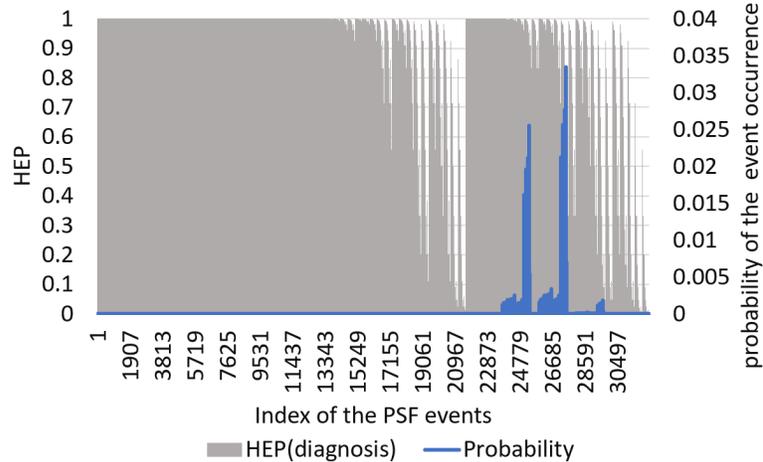


Figure 3.1: HEP distribution and joint probability of PSFs

levels along with the power-grid operating parameters. The five PSFs that are uncorrelated to the propagation of failures would remain unchanged during the cascading failures. However, the other three PSFs would change PSF levels following the power-grid state variables. Compared to [1], this approach will allow the inclusion of the detailed operator PSF levels, which in turn will allow having various HEPs, which was only four in [1].

3.3 Results

In this section, we share the results and capabilities of this model.

3.3.1 Critical PSF combinations that lead to high HEP

A striking observation from the distribution of the PSFs (adopted from Table 3 in [64]) is that not all the combinations of PSFs are equally likely to occur. For example, all the operators reported that their fitness was nominal, which implies that although there are three distinct PSF levels for fitness (nominal, degraded fitness, unfit), the probability of human operator attribute with a degraded fitness or unfit is zero. This observation is very significant in the sense that not all the 32400 events (a combination of PSF levels, each giving a HEP) have a non-zero probability of occurrence. In fact, using Table 3 of [64], we observe that from the

32400 events, only 3888 events have a non-zero probability of occurring (represents 88% of all the events). Thus, calculating HEP without considering the distribution of the PSFs can be misleading, since the PSF distributions are non-uniform. From the HEP values calculated using equation (4.2) for all the 32400 events, we can observe that 61% events have human error probability greater than 0.9. However, the joint distribution (considering independence between PSFs) of the PSFs using the PSF distributions tells that only 12% of the total events have a non-zero probability of occurrence. We plot the HEP (left vertical axis) and the probability of the event occurring (right vertical axis) with the PSF index (calculated using combinations of PSF multipliers) in Figure 3.1 Note that, to visualize the HEP and the probability of the event occurring against various combinations of PSF levels in two dimensions, we use the following equation to transform the eight-dimensional PSF levels into a two dimensional PSF index using the table 1 in [1]. The coefficient values in the following equation depends on the number of PSF levels. In total, when the maximum PSF levels are considered for all the PSFs the total number of events will add up to 32400 as mentioned above.

$$\begin{aligned}
 \text{index} = & 10800(\text{fitness}_i - 1) + 2160(\text{availabletime}_i - 1) + 432 \\
 & (\text{procedures}_i - 1) + 108(\text{ergonomics}_i - 1) + 36(\text{stress}_i - 1) \quad (3.5) \\
 & +9(\text{complexity}_i - 1) + 3(\text{experience}_i - 1) + \text{workprocess}_i
 \end{aligned}$$

Observe that the grey colored bar plot in Figure 3.1 represents HEP calculated using equation (4.2), where the PSF multiplier values are taken from Table 5.1 of [64]. Clearly, the grey bar, which represents the HEP is one for many indexes (≥ 0.9 for %61 cases), which seems exaggerated. However, the blue bar plot reveals that only a handful of those events have a non-zero probability of occurrence. The combination of HEP and distribution of PSFs can be used to identify the critical combination of PSF events that have a high probability of occurring with a high

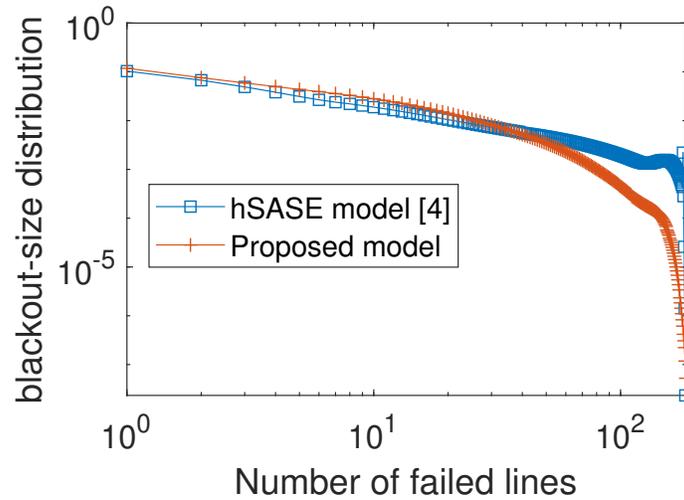


Figure 3.2: Comparison of blackout size between the proposed model and hSASE model [4].

TABLE III
CRITICAL PSF LEVELS WITH JOINT PROBABILITY > 0.01 AND HEP > 0.01

SL.	Available time		Stress		Complexity		Experience		Procedures		Ergonomics		Fitness		Work process		HEP (diagnosis)	Joint Probability
	PSF multiplier	probability																
1	1	0.55	1	0.66	2	0.67	1	0.36	1	0.81	1	0.47	1	1	1	0.69	0.02	0.023
2	1	0.55	2	0.31	2	0.67	0.5	0.53	1	0.81	1	0.47	1	1	1	0.69	0.02	0.016
3	1	0.55	2	0.31	2	0.67	1	0.36	1	0.81	1	0.47	1	1	1	0.69	0.039	0.011
4	10	0.42	1	0.66	2	0.67	0.5	0.53	1	0.81	1	0.47	1	1	1	0.69	0.092	0.026
5	1	0.55	1	0.66	2	0.67	0.5	0.53	1	0.81	10	0.17	1	1	1	0.69	0.092	0.012
6	10	0.42	1	0.66	2	0.67	1	0.36	1	0.81	1	0.47	1	1	1	0.69	0.168	0.018
7	10	0.42	2	0.31	2	0.67	0.5	0.53	1	0.81	1	0.47	1	1	1	0.69	0.168	0.012
8	1	0.55	1	0.66	2	0.67	0.5	0.53	1	0.81	50	0.36	1	1	1	0.69	0.336	0.026
9	1	0.55	1	0.66	2	0.67	1	0.36	1	0.81	50	0.36	1	1	1	0.69	0.503	0.018
10	1	0.55	2	0.31	2	0.67	0.5	0.53	1	0.81	50	0.36	1	1	1	0.69	0.503	0.012
11	10	0.42	1	0.66	2	0.67	0.5	0.53	1	0.81	50	0.36	1	1	1	0.69	0.835	0.02
12	10	0.42	1	0.66	2	0.67	1	0.36	1	0.81	50	0.36	1	1	1	0.69	0.91	0.013

HEP. We have shown an example of the critical combination of PSF events in Table 3. Here we have considered events that have $HEP \geq 0.01$ and the probability of that event occurring ≥ 0.01 . Table 3 reveals that there are only 12 events filtered using these criteria. This information is critical because it indicates that, in general, the probability of an event with high HEP is unlikely. Hence, using this approach, one can quickly identify the combinations of the PSF events (with a high HEP) that are highly likely to occur.

3.3.2 Comparison of blackout size with hSASE model [1]

We show a comparison between the blackout size distributions obtained using the proposed and the hSASE model in Figure 3.2. We simulate the Markov chain numerically for both cases with the same initial conditions ($F_i=2$, $r=0.85$, $e=0.2$, $\theta=0.2$). In our proposed model, we consider the detailed mapping of HEP with the grid variables, as shown in Table 3.2. Observe that, in Figure 3.2, the blackout size distribution in the proposed model is exponential in contrast to power-law distribution in the hSASE model. Since the hSASE model considers a coarse mapping between operator attributes and grid variables without considering the distribution of the PSFs, HEP is generally higher when the number of transmission line failures in the power grid is high. For example, in the hSASE model, the authors considered an inadequate time for the operators' when the number of transmission line failures was greater than fifty for the IEEE 118-bus system. In contrast, in the proposed model, even if there are high transmission line failures in the power grid, HEP is less than one. HEP is one only when all the transmission lines have failed, and there is nothing an operator can do. Else, operator response levels are mapped accordingly using the joint probability of the PSFs. In Figure 3.2, we have used nominal PSF multipliers for the five uncorrelated PSFs as an initial condition and then mapped HEP with grid variables. For this setting, the blue plot (hSASE) shows power-law, while the Orange plot (proposed model) shows exponential behavior. Changing the uncorrelated PSFs from nominal to high or extreme would increase the initial HEP, and hence the more drastic impact of human error is observed. This capability is not available in [1] since it does not consider the uncorrelated PSFs. This approach is more realistic because even if there are a high number of transmission line failures due to cascading, grid operators often successfully save the remaining lines by implementing effective

strategies. Thus, the proposed work adds more fidelity compared to the hSASE model.

Chapter 4

Balancing Smart Grid's Performance Enhancement and Resilience to Cyber Threat

The strong interplay between the power grid and the corresponding communication and control network plays a pivotal role in the resilience of the smart grid. In this work, the dynamics of the interdependence among smart-grid subsystems such as the power grid, communication network, and response of human operators are captured during the propagation of cascading failures. A previously developed Markov-chain based model is refined into an interdependent Markov chain model to capture the role of cyber threat from the communication network and the human-operator error during cascading failures. The state transitions of the Markov chain are parameterized by the critical operating parameters of the power grid. The calculations assume a generic form of correlation between the level of and damage from cyber-attacks, on the one hand, and the level of interdependence on the other hand. The model finds the optimal level of interdependence, i.e., the trade-off between well-informed control and vulnerability to attacks that minimizes the probability of massive cascading failures in power grids. There is a point of diminishing return beyond which the harm of exposure to cyber threat outweighs the benefits of information.

A schematic of an interdependent smart-grid system that integrates the power-grid, the communication network, and the human-operators, as well as cyber attackers in the loop is shown in Figure 4.1 (with details of each variable and parameters described subsequently). In this chapter, we refine a previously developed single-layer Markov-chain based analytical cascading failure model to an interdependent Markov chain model. We term it as the model for Cascading Failures in

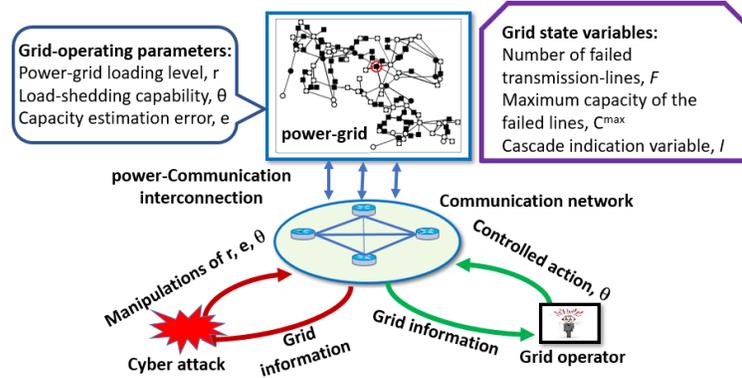


Figure 4.1: An interdependent smart grid that integrates the power-grid, the communication network (with cyber attackers), and the grid operators.

Interdependent Stochastic Abstract State-space Evolution (I-SASE). The I-SASE model can capture the benefits of having more information through the capability of implementing load shedding during contingencies. However, with more power-communication interdependence, the grid becomes vulnerable to cyber-attacks through the communication channels. We use the model to find the optimal interdependence between the power grid and communication network that minimizes the risk of massive blackouts by taking the potential risk of a cyber-attack into account.

4.1 I-SASE Analytic Framework

The I-SASE model analytic framework depends on the three layers of the smart grid. In this section, we discuss the role of each layer on the dynamics of cascading failures.

4.1.1 Influence of the communication network on the propagation of cascading failures

Figure 4.2 shows the interdependence between the power grid and the communication network used in our model. We consider that the power grid and the communication network can have different topologies; however, one power node can be connected with one communication node to share the grid information to

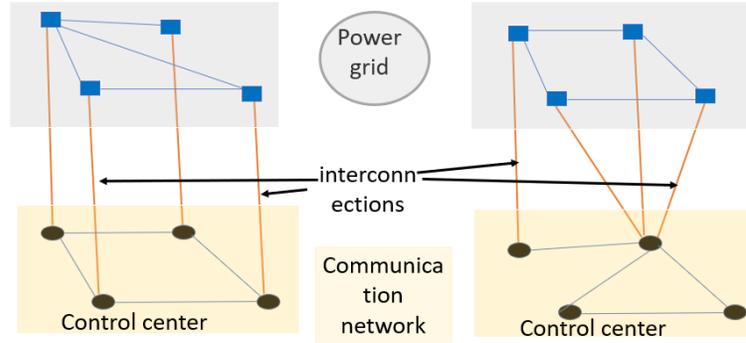


Figure 4.2: A schematic diagram showing the inter-connectivity between the power grid and the communication network in smart grids

the control center, i.e., if the power grid has N nodes (buses) then we can have a maximum N power-communication interconnections, N_{pc}^{max} . Increased interdependency through the interconnections gives additional information and hence reliability through informed control. Power-communication interconnections, N_{pc} , aids human operators to actuate system-wide control actions to mitigate the risk and spread of cascading failures [7]. However, interconnections come at the price of higher infrastructure cost as well as the cost to minimize the risk of cyber threat through the interconnections. Thus, the security risk of the system can be minimized by integrating better security policies that require higher costs. In real-world scenarios, SCADA networks connect many, but not all, nodes in a power network over communication networks using fiber-optic, microwave, and telephone communication channels combined [7]. Hence, in practice, $N_{pc} \leq N_{pc}^{max}$.

Note that Figure 4.2 (left) and 4.2 (right) show four interconnections between the power grid and communication network in two different ways. With this connectivity structure at hand, we discuss the role of the communication network during the propagation of failures. The ratio of power-communication interconnections and maximum number of power-communication interconnections is $k = N_{pc}/N_{pc}^{max} \in [0, 1]$. A higher value of k indicates strong interactions between the power grid and communication network. While some of the similar

interdependent models consider that failure in a power node initiates failure in the communication network with some probability [44], in this chapter, we do not consider failures in the communication network directly rather we capture the influence of disturbance in power grid on communication network indirectly. We consider that transmission line failures in the power grid can trigger the corresponding power node (bus) failure and when all the transmission lines that are connected to a power node fail, we consider the corresponding power node as failed. When a power node fails, we assume that the corresponding interconnection has also failed. Thus, with power node failures, less information on the grid will be available to the control operators. Here we assume that the intra-interactions inside a communication network is uncorrelated with any incident in the power grid. The communication nodes are protected using back up power provided through batteries. Thus, failures in the power node do not necessarily initiate failures in the communication node. Note that the study of intra-interactions inside a communication network (see Figure 4.2), failures in communication nodes, and their influence on the power grid are beyond the scope of this work.

Similar to [5], we consider $\theta \in [0,1]$ as the load-shedding constraint, i.e., the ratio of uncontrollable loads (loads at which load shedding cannot be performed) and the total load in the power grid. In [5], the authors consider θ as a fixed parameter. However, in this work, we consider θ as a dynamic parameter that is inversely correlated with k , i.e., when $k = 1$, $\theta = 0$ and when $k = 0$, $\theta = 1$. This indicates that, when $N_{pc} = N_{pc}^{max}$, there are no load-shedding constraint. On the contrary, when $N_{pc} = 0$, we cannot implement any load-shedding in the power grid. Load-shedding constraint increases with failures of power-communication interconnections since communication sensors cannot send measurement data and control signals cannot be sent through the failed power-communication interconnections. To capture the dynamics of load-shedding constraints during cascading

failures, we define θ as

$$\theta_i = \theta_{ini} + \frac{klF_i}{N_{pc}^{max}}, \quad (4.1)$$

where $\theta_{ini} = 1-k$. Here F_i is the number of failed transmission line in the power grid due to contingency, and $l = N/M$, where N is the number of power nodes (buses) and M is the number of transmission lines. At this point, intuitively we can correlate that increasing power-communication interconnections causes θ_{ini} to decrease. Thus, having a maximum level of power-communication interconnections would increase flexibility to implement necessary control actions by the power grid operators during contingencies.

However, power-communication interconnections couple the power grid and communication network into a single large network, which causes the network to be susceptible to attacks. The probability of human operator errors due to additional information processing, the vulnerability of the communication network, noise in the communication channel, probability of intentional cyber-attacks increases with the number of N_{pc} . This assumption is very intuitive and yet realistic since having more interconnection increases the likelihood of exposure to the entire network. To capture this phenomenon, we introduce a cybersecurity threat parameter, ψ , which captures the vulnerability of a single large network. The parameter ψ depends on the number of interconnections (N_{pc}), and a large number of interconnections would imply more opportunities for attackers to breach the security of the data. Therefore, the parameter ψ can be modeled reasonably as

$$\psi_i = \lambda k \left(1 - \frac{\exp(\gamma l F_i) - 1}{\exp(\gamma N_{pc}^{max}) - 1} \right), \quad (4.2)$$

where $\gamma \in [0,1]$ is a constant that shapes the exponential function, and $\lambda \in [0,1]$ is a constant used to capture the probability of cyber threat due to added

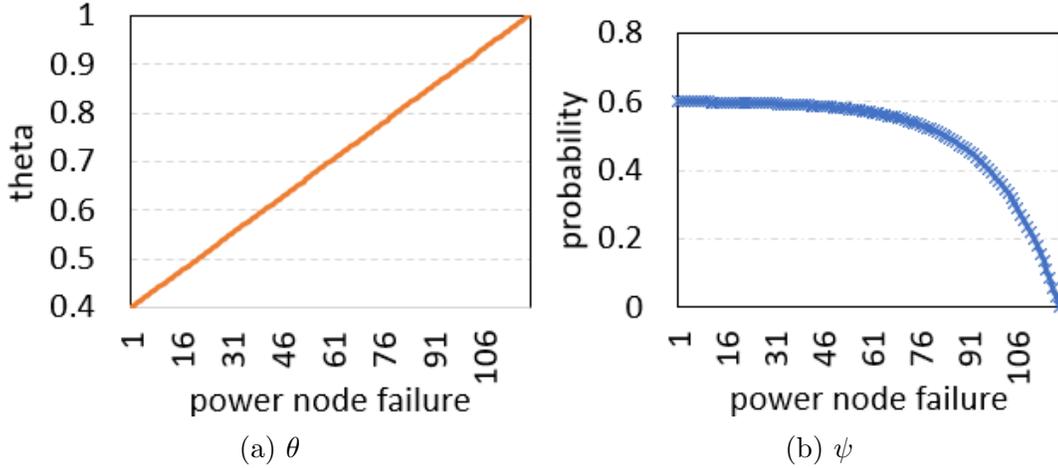


Figure 4.3: (a) Changes in load shedding constraint and (b) the probability of cyber threat due to power failures and $k=0.6$.

interdependence between power and communication network. Note that when there are no failures in the power grid, $\psi_i = \lambda k$, i.e., the probability of cyber threat depends on the N_{pc} and λ . When failures propagate in the power grid during a cascade, the probability of a cyber threat decreases as the size of the power grid decreases due to failures in the power nodes and N_{pc} . Note that the assumed linear and exponential relationships for θ_i and ψ_i can be any arbitrary monotone function, respectively.

4.1.2 Influence of the human operators' on the propagation of cascading failures

Human operators play a pivotal role in mitigating the propagation of failures. During everyday operation, operators/engineers need to tackle various contingencies to improve the reliability of the grid. A key part of the job of a power grid operator is working under stress and time constraint. In most cases, power grid operators can implement control mechanisms according to the prescribed plan to minimize the effect of any contingencies. However, human operator error during propagation of failures can increase the probability of a large blackout exponentially. Wang *et al.* introduced a Markov chain framework to integrate the proba-

bility of human errors into the state-space of the Markov chain transition matrix. Specifically, the authors [1] quantize human operator errors in four distinguishable levels and use the state variables of the Markov chain to define the transition between human operator error levels. This work used a previous work [64] by Abreu *et al.*, where they used eight performance shaping factors (PSFs) to define the attributes of human operation and used the Standardized Plant Analysis Risk - Human (SPAR-H) methodology to calculate the human error probability analytically. Abreu *et al.* calculated the probability distribution of the PSFs empirically using grid operator interviews.

In this chapter, we use an earlier reported work from chapter 3 that represents the role of operators' performance shaping factors (PSFs), and human-error probability (HEP) as a function of the smart grid operating conditions, such as available time, complexity, and human performance attributes, e.g., stress, training, and working environment. Specifically, the Markov-chains transition matrix includes the dynamics of the smart-grid operator attributes through the HEP. The mapping between grid conditions and operator response levels utilizes the probability distribution of the PSF levels and is established based on a histogram-equalization principle, assuming a monotone relationship between the HEP values and number of line failures. The work allows one to calculate the HEP for each grid state of the Markov chain conditional on the PSFs (see [87] for details).

4.1.3 State variables of the Markov chain

Similar to the SASE model [5], the state space of the I-SASE model is defined as $S_i = (F_i, C_i^{max}, I_i)$; where F_i, C_i^{max}, I_i are the states variables of the Markov chain. Here, F_i is the number of transmission-line failures at state S_i . To be consistent with [5], we also allow one transmission-line failure in our Markov chain at a time, i.e., during cascading failure we consider a minimal time interval such that only one new failure can occur in that period. The variable C_i^{max} represents

the maximum capacity of the failed transmission lines at the current state, and I_i is the cascade stability indicator. In this work, we also use the same set of five capacities used in [5]. In addition to the load shedding constraint (θ) parameter, we use two additional parameters; the power grid loading level, r , which is defined as the ratio of the total load demand and the generation capacity of the power grid [5] and the capacity-estimation error, e , which represents the error (at the control center) in estimating the actual capacity of the transmission lines. A higher value of r is indicative of stress on the power grid from [5]. In our Markov-chain model, parameter e is used to determine overloaded lines, a parameter that naturally affects cascading failures profoundly.

4.1.4 Probability transition matrix

The transition matrix of the Markov chain is a $2M|\mathcal{C}| \times 2M|\mathcal{C}|$ matrix, where M is the total number of failed transmission lines and $|\mathcal{C}|$ is the cardinality of the set of capacities. In our model, the number of transmission-line failure, F_i , increases monotonically with one failure per unit time since no healing of the nodes/lines were considered. Hence, the transition matrix \mathbb{P} is a row-stochastic upper diagonal matrix. The state transition probability, $f(S_j|S_i)$, from state $S_i = (F_i, C_i^{max}, I_i)$ to the next state $S_j = (F_j, C_j^{max}, I_j)$, is defined below:

$$f(S_j|S_i) = \begin{cases} 1 & \text{if } F_j = F_i, C_j^{max} = C_i^{max}, I_j = I_i = 1, \\ P_{stop}(S_i)(1 - \psi_i)(1 - h_i) & \\ \text{if } F_j = F_i, C_j^{max} = C_i^{max}, I_j = 1, & \\ P(S_j|S_i) & \text{if } F_j = F_i + 1, C_j^{max} \in \mathcal{C}, I_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4.3)$$

Here $P_{stop}(S_i)$ is the probability of cascade-stop at state S_i , and $P(S_j|S_i)$ is

the probability of a new transmission-line failure. The parameters ψ_i and h_i capture the role of cyber-security threat and human operator error, respectively. Cascading failures stop when the Markov chain reaches to an absorbing state, i.e., the chain reaches a terminal state with $P_{stop}(S_i) = 1$. Note that $P_{stop}(S_i)$ is the key parameter that dominates the dynamics of I-SASE model and the characterization of $P_{stop}(S_i)$ was discussed extensively in [5].

Following [5], $P_{stop}(S_i)$ is defined as a weighted combination of $P_{stop}(F_i)$ (probability of cascade-stop with F_i failures) and $P_{stop}(C_i^{max})$ (probability of cascade-stop with capacity C_i) as follows,

$$P_{stop}(S_i) = wP_{stop}(F_i) + (1 - w)P_{stop}(C_i^{max}), \quad (4.4)$$

where we use $w = 0.5$. Note that $P_{stop}(F_i)$ has a bowl-shape pattern with three distinguishable phases each defining three phases of a cascading failure as [5]

$$P_{stop}(F_i) = \begin{cases} a_1\left(\frac{a_2M - F_i}{a_2M}\right) + \epsilon & \text{if } 1 \leq F_i \leq a_2M, \\ \epsilon & \text{if } a_2M < F_i \leq 0.6M, \\ Q(F_i) & \text{if } 0.6M < F_i \leq M, \end{cases} \quad (4.5)$$

where $Q(F_i)$ is a fixed quadratic function approximating the tail of the family of bowl-shape functions (see Figure 4 in [5]). Moreover, $P_{stop}(M) = 1$, i.e., there are no more transmission-line failures if $F_i = M$. Following the formulation shown in [1], a reasonable parametric model for the scaling parameters a_1 and a_2 and ϵ is

$$a_1 = \max(0.02, 0.4 - 0.25[r]_{0.5} - [e]_0^{0.5}(0.2 - [e]_{0.1}^{0.5}) - 0.25\theta_i) \quad (4.6)$$

$$a_2 = \max(0.01, 0.1 - 0.05[r]_{0.5} - 0.1[e]_{0.1}^{0.5}(0.2 - [e]_0^{0.5}) - 0.07\theta_i) \quad (4.7)$$

$$\epsilon = \max(0.01, 0.6 - 0.4[r]_{0.5} - 0.5[e]_0^{0.5} - 0.3\theta_i). \quad (4.8)$$

Similarly, power-grid simulations show that $P_{stop}(C_i)$ is high (low) when the capacity of the failed transmission line is low (high). Hence, $P_{stop}(C_i)$ is defined as

$$P_{stop}(C_i) = \max \left\{ a_3 \left(\frac{\max\{\mathcal{C}\} - C_i^{max}}{\max\{\mathcal{C}\}} \right)^4, a_4 \right\}. \quad (4.9)$$

Following the power grid simulations described in [5, 1], a reasonable parametric model for the scaling parameters a_3 and a_4 is

$$a_3 = \max(0.02, 0.4 - 0.2[r]_{0.5} - [e]_0^{0.5}(0.2 - [e]_{0.1}^{0.5}) - 0.3\theta_i) \quad (4.10)$$

$$a_4 = \max(0.01, 0.1 - 0.06[r]_{0.5} - 0.1[e]_0^{0.5}(0.2 - [e]_0^{0.5}) - 0.06[\theta_i]) \quad (4.11)$$

Note that, the constants in the scaling parameters a_1 , a_2 , ϵ , a_3 , and a_4 are empirically calculated based on the simulation results (part of the results have been shown in [6, 5, 1]).

Transition probabilities, $P(S_j|S_i)$ of the Markov chain, are calculated as follows:

$$P(S_j|S_i) = \begin{cases} P_{cont}(S_i)(1 - P_{hc}(S_i)) & \text{if } C_j^{max} = C_i^{max}, \\ P_{cont}(S_i)P_{hc}(S_i) \frac{w(C_j^{max})}{\sum_{m: C_m > C_i^{max}} w(C_m)} & \\ \text{if } C_j^{max} > C_i^{max}, \end{cases} \quad (4.12)$$

where $P_{cont}(S_i) = 1 - P_{stop}(S_i)(1 - \psi_i)(1 - h_i)$ and

$$P_{hc}(S_i) = \min(1, \alpha(F_i + \beta)^3), \quad (4.13)$$

where α and β are constants. Moreover, $P_{hc}(S_i)$ is the transition probability to a state with higher capacity and the $w(C_k)$'s are capacity weight parameters calculated from power grid simulations in [5]. We adopt the formulation of the

human-error probability from [87] and ψ_i using (4.2), and after each failure in the power grid, we update the parameter θ_i parameter using (4.1).

4.2 Results

The parametric model in [5] is tuned to the IEEE 118-bus topology. In the following calculations, we use the same topology with 118 power nodes and 186 transmission lines.

4.2.1 $P_{stop}(S_i)$ and blackout size distribution for various k

Cascade-stop probabilities (for $k=0.1$, $k=0.5$ and $k=0.9$) and the distribution of blackout size (for $k=0.1$, $k=0.5$, and $k=0.9$) as a function of the number of failed transmission lines are shown in Figure 4.4 and 4.5, respectively. Note that in Figure 4.5 the hump at the tail of the distribution for $k=0.9$ indicates the heavy tail nature of the distribution for high power-communication inter-connectivity. The cascade-stop probabilities were calculated for the same initial conditions with two transmission line failures by simulating the Markov chain. From Figures 4.4 and 4.5, it can be observed that for $k=0.5$, the cascade stop probability is higher compared to the other two cases, and the blackout size distribution is less severe when $k=0.5$. This indicates that when power-communication interdependency is either very high or very low, there is an added probability of a large blackout.

4.2.2 Optimal power-communication interdependency

From the I-SASE model formulation, it can be observed that with transmission line failures in the power grid, the parameters θ and h increases, and ψ decreases. This is intuitive but informative. Failures in the power grid reduce the capability of implementing load-shedding during cascade propagation. Also, failures in the power grid increase the probability of human-operator error. On the other hand, failure in the power grid reduces the communication-power interdependency since the corresponding power-communication interconnection also fails. Due to this in-

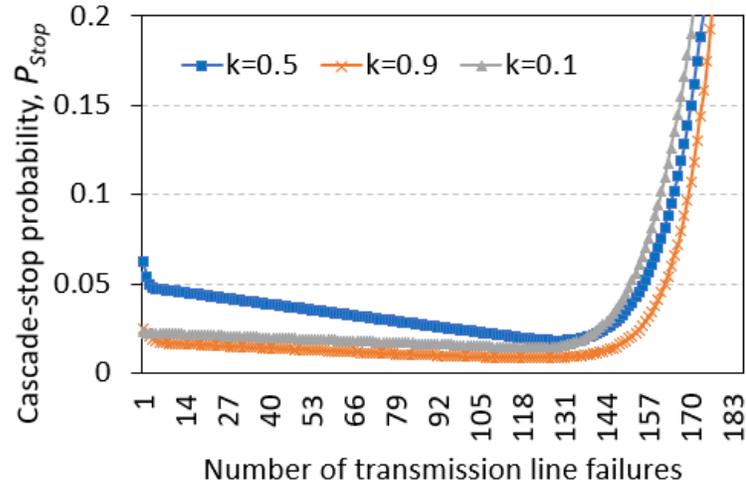


Figure 4.4: Cascade stop probability for $k = 0.5$ and 0.9 . For $k = 0.5$, P_{stop} is high initially compared to $k= 0.9$, which indicates that the probability of a large cascade for $k = 0.5$ is lower. To visualize the effect of cascade stop probability, the y -axis is truncated from $[0,1]$ to $[0,0.2]$

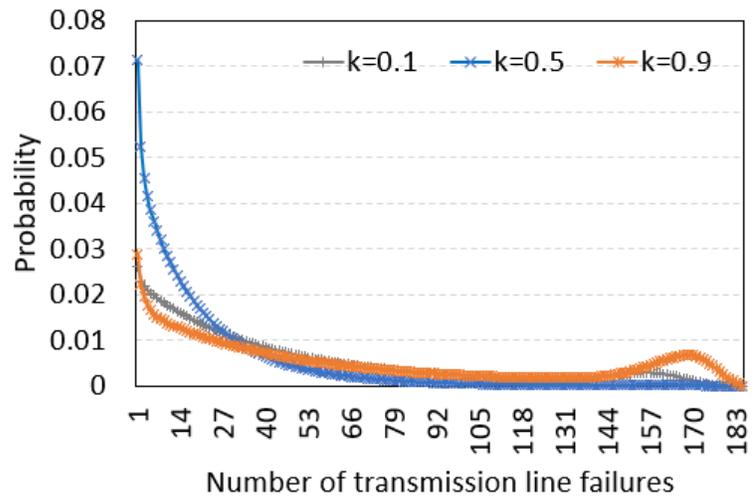


Figure 4.5: Distribution of the blackout size for $k = 0.1$, $k = 0.5$ and $k = 0.9$. For $k = 0.1$ and $k = 0.9$, blackout size distribution shows heavy tail, which indicates a power law distribution as compared to an exponential distribution for $k = 0.5$.

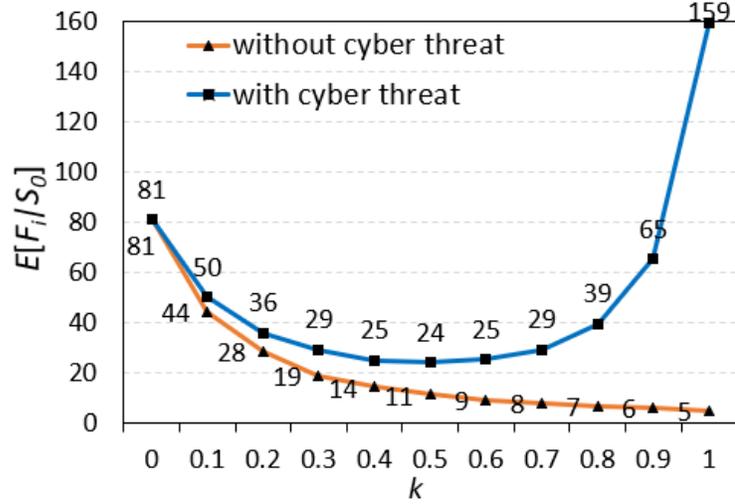


Figure 4.6: Optimal power-communication interdependency with and without cyber threat. Without cyber threat, $E[F_i|S_0]$ increases with k with minimum at $k = 1$. Including the cyber-security parameter, ψ_i , $E[F_i|S_0]$ is minimum when $k = 0.5$.

verse nature, there exists an optimal power-communication interdependency that minimizes the expected number of transmission line failures given an initial condition, $E[F_i|S_0]$ subject to $0 \leq \theta, \psi, h \leq 1$ and for a given grid condition during cascade initiation. In Figure 4.6, we plot the average transmission line failures for various k with and without considering the effect of cyber threat. Here we consider $r = 0.6$ and $e = 0.1$ for all the cases. It is clearly visible from Figure 4.6 that the optimal power-communication interdependency is when $k = 0.5$ (24 line failures in the square marked blue line). However, when we do not consider the cyber-threat effect, the average transmission line failures were minimum (5 in the triangle marked orange line) when the power-communication interdependency is maximum, i.e., $k = 1$, which was also reported in [7].

A comparison of the effect of various levels of power-communication interdependency on cascading failures between [7] and I-SASE model is shown in Figure 4.7. Figure 4.7 (a) is adopted from [7] and 4.7 (b) using the I-SASE model. the Y-axis in Figure 4.7 (a) represents the % of loads served after the cascade ends.

f represents various % of initial line failures. The ideal case (represented by the black line) represents scenario when communication failures does not effect power failures. The vulnerable case (represented by the blue line) represents the work in [21], where failure in power has deterministic impact on communication. The intermediate case (represented by the red line) represents the work by Korkali *et al.* where failures in communication fails power nodes probabilistically. If we follow the red line in 4.7 (a), we can see that as the interdependency is increased, the % of loads served after the cascade ends increases and maximum is found when the interdependence is one. In Figure 4.7 (b) we show the results obtained using the I-SASE model. We assume a linear correlation between the % of average survived lines with % of load served. If we don't consider cyber threat (represented by the orange line), we also obtain the similar result as reported in [7]. However, our model has the additional capability of capturing the harm of excessive communication and cyber threat. If cyber threat is considered, then we see that the % of average survived lines increases initially (represented by the grey line) with increase in interdependence, however, there exists a point of diminishing return beyond which the harm of communication outweighs the benefits of communication.

The knowledge of this trade-off between benefits and risks of having more information (optimal power-communication interdependence) can be used as a useful design parameter to mitigate cascading failures in the smart grid.

4.2.3 Role of operators performance on the propagation of cascading failures

Note that in the transition matrix of the I-SASE model, the role of human-operator error is captured through parameter h_i (multiplying h_i with $P_{stop}(S_i)$). The model in [87] correlates the human error probability with the states of the Markov chain, i.e., for every transitory state of the Markov chain, we have an

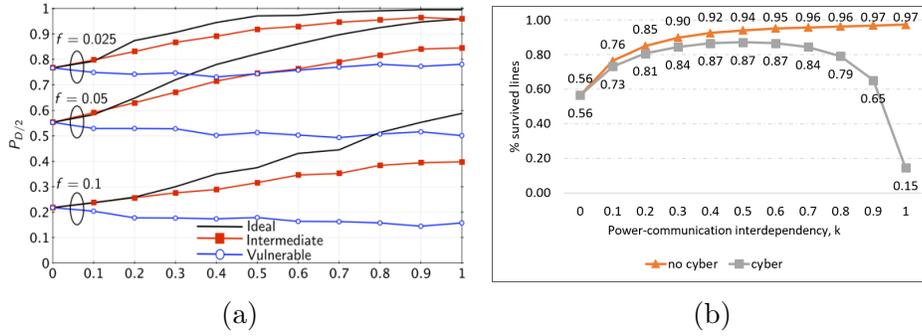


Figure 4.7: Comparison of the effect of various levels of power-communication interdependency reported in [7] and I-SASE model (a) % of served loads after cascade ends for various interdependency levels (adopted from [7]) (b) % of survived lines for various interdependency levels using I-SASE model

associated human error probability which is governed by the PSFs of the operator. Failures in the power grid increase HEP monotonically, and the formulation is described extensively in [87]. In Figure 4.8, we show the role of human operator error on cascading failures. We simulate the Markov chain considering two line failures and the same initial values of the r, e, k parameters as in the previous section. In Figure 4.8, we show two plots, one with extreme PSF levels (orange) and the other with nominal PSF levels (blue). Note that although the other parameters of the Markov chain are same, various initial PSF levels can exist due to the different level of experience of the operator, ergonomics, work process, operating procedure availability, and so forth. In addition, during the propagation of failures, PSFs such as available time, stress, and complexity changes, which make the HEP dynamic and state dependant. Observe from the distribution of the blackout size in Figure 4.8 that, for nominal PSFs (24 lines failures on average), the effect of cascading failures is less severe compared to the extreme PSFs (32 lines failures on average).

4.3 Cyber threat and operator error aware stochastic model for cascading failures

In this section, we refine the I-SASE model further, which captures the benefits and risk of information through the communication network. Specifically, we

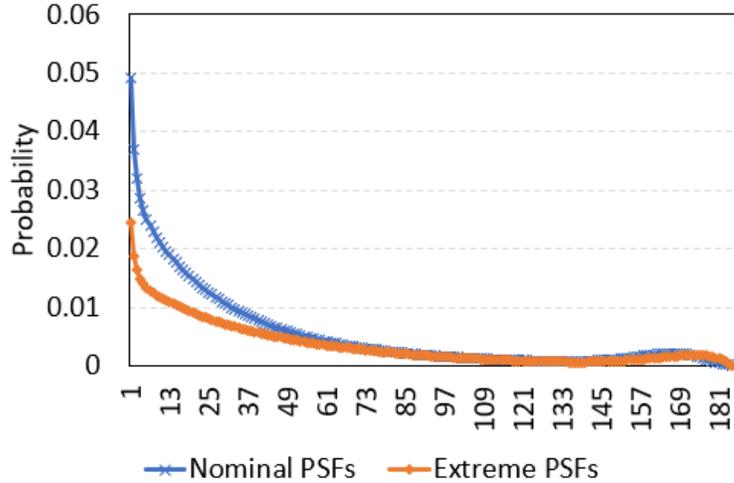


Figure 4.8: Distribution of the blackout size for Nominal and extreme HEP and $k=0.5$, $F_i=2$. $E[F_i|S_0]$ for nominal and extreme HEPs are 24 and 32 lines respectively.

introduce the cyber threat variable as a state variable of the Markov chain to capture the detailed dynamics of the affect of cyber threat.

4.3.1 State variables of the Markov chain

We consider the following three state variables of the Markov chain: F_i , the number of transmission line failures at state S_i ; C_i^{max} , the maximum capacity of the failed transmission liens at state S_i ; Ψ_i , the level of cyber threat in the communication network, and I_i , a binary variable indicating whether the state is an absorbing or transitory state in a Markov chain. Thus the state space of the model is defined using $S_i = (F_i, C_i^{max}, \Psi_i, I_i)$. Index of a state is defined using $2|\mathcal{C}||\Psi|(F_i - 1) + 2|\Psi|C_i^{max} + 2(\Psi_i - 1) + I_i + 1$, where $|\mathcal{C}|$ is the cardinality of the set of capacities and $|\Psi|$ is the cardinality of the set of cyber threat models. For example, in IEEE 118- bus network, the total number of transmission line, $M = 186$, considering $|\mathcal{C}| = 5$ and $|\Psi| = 5$, there are 5580 distinct states of the Markov chain. Similar to [5], we consider one failure transitions in the Markov chain, i.e., the overall duration of the cascading failures is divided in such a short time (ΔT) such that only one failure is allowed. In addition to the state variables,

three power grid parameters are used to define the transition probabilities of the Markov chain. We also consider the three parameters used in [5] to model the state transitions, the ratio between the load and the maximum generation in the grid, r ; the uncertainty in the flow of information, i.e., the error in estimating the power flow in transmission lines namely capacity-estimation error, e and the constraint over implementing load shedding in the grid, θ . However, we also consider the dynamics of system operator error and a cyber threat for modeling the state transitions, which is not considered in [5]. In that sense, our model is a generalization of [5], and if we consider fixed parameter values for r, e, θ and do not consider cyber threat and operators' error, our model collapses to the original SASE model.

4.3.2 Effect of interdependency among the various layers of the power grid

Interdependence between power grid and the associated communication network can lead to three scenarios contributing cascading failures: (1) effective handling of information can minimize cascading failures, (2) increased communication capability also adds vulnerability to the grid in the form of cyber-attacks and (3) the incorrect actions of the system operators dealing with cascading events can lead to further escalation of the event. In the following subsections, we describe the interactions among various layers of the power grid and model their influence on cascading failures.

The information on the grid is conveyed to the control center through the communication network infrastructure. The topology of the grid and the communication network are not necessarily identical. However, power grid nodes (i.e., substations) are equipped with communication facilities to send (receive) information (instructions). Not necessarily all the grid nodes are connected individually to communication networks. Supervisory control and data acquisition, a com-

puter system (SCADA) networks connect power and communication nodes using different mediums (e.g., fiber, microwave) and use this interconnection to collect information of all the nodes of the grid [7]. Figure 4.12 shows two such examples of inter-connectivity between the power grid and communication network. With this setting, we consider maximum inter-connectivity when all the power grid has an associated interconnection, i.e., if the grid has N nodes (buses), then there can be a maximum of N power-communication interconnections, N_{pc}^{max} . Power-grid system operators use these interconnections, N_{pc} , to send appropriate control actions to maintain stable health of the grid [7]. Intuitively, the maximum number of interconnections should maximize the availability of all the information. However, in practice, due to constraints on communication infrastructure cost and cost of hardening the communication network from cyber threats, $N_{pc} \leq N_{pc}^{max}$. We refer level of power-communication inter-connectivity, k as the ratio between power-communication interconnection to its maximum, i.e., $k = N_{pc}/N_{pc}^{max} \in [0, 1]$, where a higher value to k indicates stronger inter-connectivity. Failure of interconnection is considered as loss of information. While some literature considers a direct effect of failures in the communication network on the power grid [44, 7] during cascading events, others argue that the effect of failures in the communication network is independent of failures in power grid and are uncorrelated. In this work, we consider that loss of information during cascading failures is initiated from power node failures. Communication nodes are equipped with backup power through batteries. Since cascading failures happen in a short duration in time, the probability of a communication node failure due to a transmission line/ power node failure is highly unlikely, i.e., the intra-connectivity within the communication network is independent of any incidents in the power grid. Transmission line failures increase the probability of a power node (bus) failure. When all the transmission lines associated with power nodes fail, we consider the failure of a

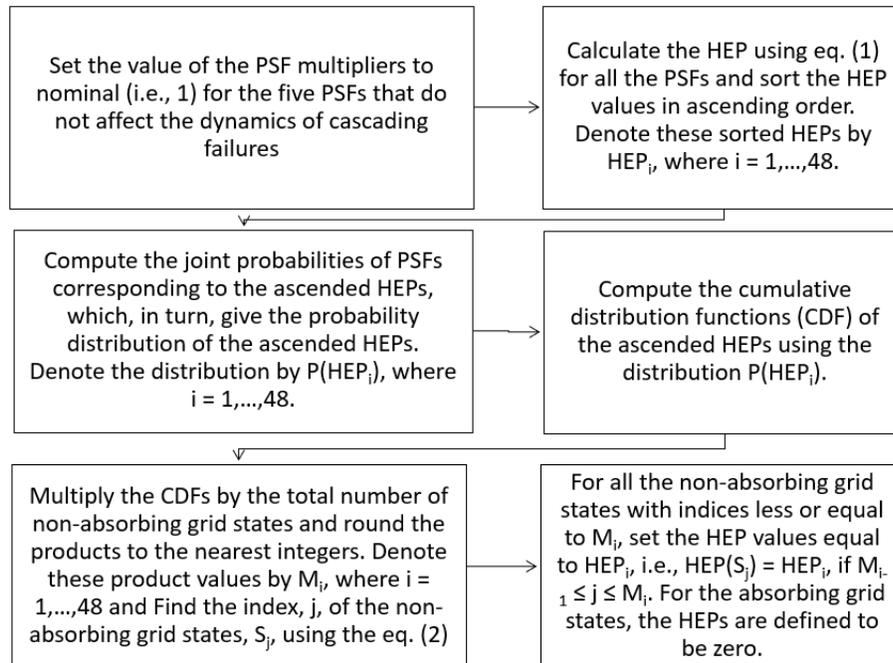


Figure 4.9: A flowchart for mapping HEP with the state variables of the Markov chain

power-communication interconnection.

System operators influence on the reliability of power grids

From preventive maintenance to protective actions, system operators are actively involved in managing the health of the grid and mitigating any risk that can lead to an outage scenario. However, errors in the decision during contingencies by the system operators can lead to a larger blackout than usual. System operators have to deal with contingencies under time constraints, stress, and the complexity of the problem varies significantly. Again, factors such as experience, work process, procedures, ergonomics, fitness can affect the capability of making an appropriate decision during contingencies. Following the work in [64, 1], in earlier work, we have developed a model to map the operators' error probability with the grid states of the Markov chain [88]. First, the model finds the human-error probability (HEP) as a function of operators' performance shaping factors (PSFs), and the grid operating conditions, such as available time, complexity, and human

performance attributes, e.g., stress, training, and working environment. Then, a mapping between grid conditions and operator response levels is established utilizing the probability distribution of the PSF levels. The mapping is established based on a histogram-equalization principle, assuming a monotone relationship between the HEP values and the number of line failures. We show a summarized flowchart of the steps used to map the operators' error probability with the state variables of the Markov chain in Figure 4.9 (discussed in chapter 3). Three realizations (representing nominal, medium, and extreme PSF values from [64]) of the system operator's error probability against the indexes of the state are plotted in Figure 4.10. Note that for nominal PSFs, the human error probability is mostly zero as opposed to one for extreme along with the index of the state. For the medium PSFs, the error probability monotonically increases to one with the state index. Depending on the independent PSF factors, the human error probability plot varies between nominal and extreme scenarios along with the state index. Note that Wang *et al.* used a separate state variable for human factors [1]. The benefit of using the proposed mapping is that the operator's error probability can be calculated as a function of the state variables of the Markov chain. So, the need for a human factor state variable is redundant, which ensures a significant reduction in the state-space of the transition matrix. The operators' error probability affects the transition probabilities of the Markov chain through the load shedding constraint.

Influence of cyber threat on the reliability of the power grid

A major enhancement of our proposed model compared to the existing model is its capability to capture the dynamics of cyber threats. The single giant network made of the power grid and the associated communication network makes the entire network susceptible to cyber attacks. Attackers can use the communication network as a medium to get into the grid network and eventually initiate cascad-

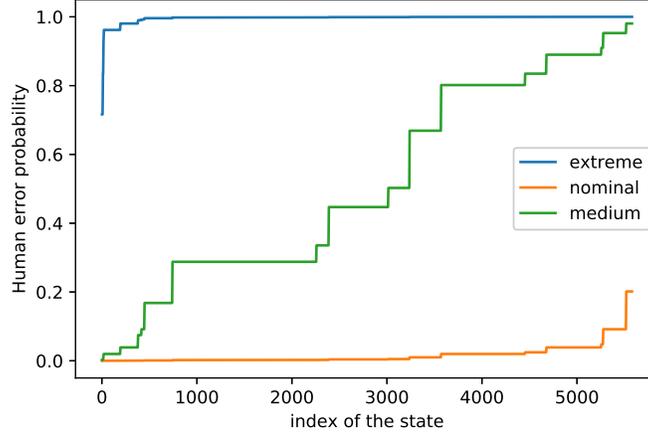


Figure 4.10: Human error probability mapped with state indexes of the Markov chain for various operators' PSFs

ing failures by manipulating data, false data injection, and changing the system parameters. Increasing k would increase communication capability in the grid but, at the same time, increase the risk of cyberattacks. To capture this critical attribute of the grid, we introduce the state variable Ψ , which can be categorized to indicate the various cyber threat profile for the grid. For example, we consider three distinct cyber threat level, $\Psi \in [\Psi_1, \Psi_2, \Psi_3]$ in this work to capture the scenarios from low, medium and high cyber risk. Monotone cyber transitions are considered in this work, i.e., with more transmission line failures, cyber transitions from low to high are considered not vice versa.

4.3.3 Transition matrix of the Markov chain

The transition matrix of the Markov-chain developed following the approach in [5, 1]. The state-space of the Markov chain is a $2M|\mathcal{C}||\Psi| \times 2M|\mathcal{C}||\Psi|$ matrix. In the subsequent calculations we consider $|\mathcal{C}| = 5$ and $|\Psi| = 3$ i.e., five types of transmission line capacities and three levels of cyber threat (low, medium and high). In our model, we did not consider any healing of the failed lines during the propagation of cascading failures, i.e., the number of failed lines increases

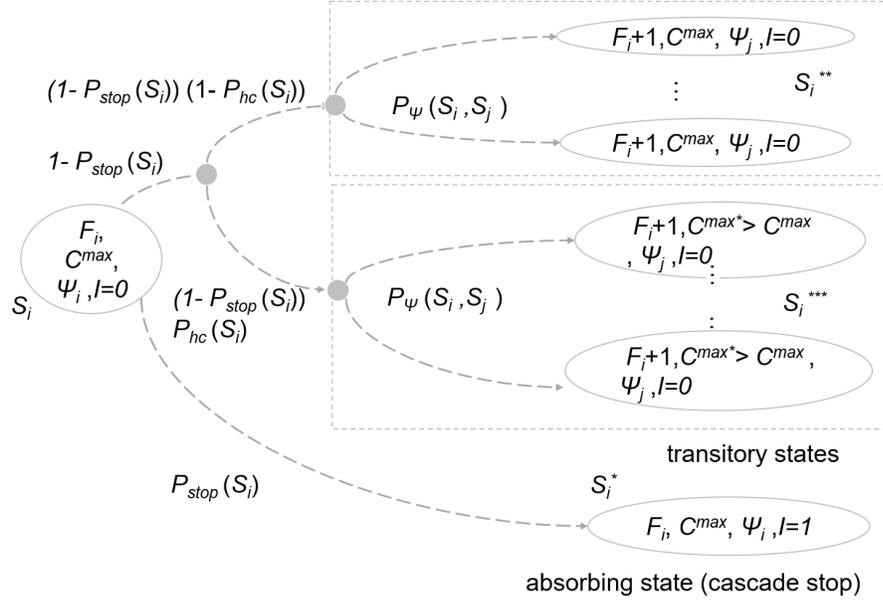


Figure 4.11: Schematic diagram of the state transitions of the Markov chain

monotonically until it reaches to an absorbing state. There are three types of transitions allowed in the Markov chain, from a transitory state to another transitory state with the same capacity, higher capacity, or an absorbing state. We show the various types of transitions of the Markov chain using figure 4.11. When an absorbing state is reached, cascading failures end for that iteration. For transitioning between transitory states, we allow one additional failure per unit time, i.e., from F_i to $F_i + 1$ there are $|\mathcal{C}||\Psi|$ possible transitions within the Markov chain. Due to the monotonic assumption of the line failures, the transition matrix \mathbb{P} is a row-stochastic upper diagonal matrix. The transition probabilities from state $S_i = (F_i, C_i^{max}, \Psi_i, I_i)$ to state $S_j = (F_j, C_j^{max}, \Psi_j, I_j)$, is given below:

$$f(S_j|S_i) = \begin{cases} 1 & \text{if } F_j = F_i, C_j^{max} = C_i^{max}, \Psi_j = \Psi_i, \\ & I_j = I_i = 1, \\ & P_{stop}(S_i) \\ & \text{if } F_j = F_i, C_j^{max} = C_i^{max}, \Psi_j = \Psi_i, \\ & I_j = 1, I_i = 0, \\ P(S_j|S_i) & \text{if } F_j = F_i + 1, C_j^{max} \in \mathcal{C}, \\ & \Psi_j \in \Psi, I_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4.14)$$

Here $P_{stop}(S_i)$ and $P(S_j|S_i)$ represents the probability of cascade stop at state S_i and cascade to continue at state S_j from S_i respectively and M is the total number of transmission lines in the grid. We refine $P_{stop}(S_i)$ from [5] as follows to include the effect of cyber threat:

$$P_{stop}(S_i) = (wP_{stop}(F_i) + (1 - w)P_{stop}(C_i^{max})) (1 - k\phi_i(\Psi_i)) \quad (4.15)$$

Here $\phi_i(\Psi_i)$ represents cyber threat probability for any given state and is defined as follows:

$$\phi_i(\Psi_i) = \lambda k \left(1 - \frac{\exp(\gamma l F_i) - 1}{\exp(\gamma N_{pc}^{max}) - 1} \right), \quad (4.16)$$

Note that the term $1 - k\phi_i(\Psi_i)$ is high when k is low indicating cyber threat has low effect on cascade stop probability and cascading failures. On the contrary, When k is high, $1 - k\phi_i(\Psi_i)$ is low, which indicates that higher inter-connectivity yields higher probability of cyber threat that reduces the probability of cascade stopping. Again, in both cases, $1 - k\phi_i(\Psi_i)$ decreases from k to zero for zero line failures to complete blackout respectively indicating that line failures fails power-communication inter-connectivity, which in turn reduces the cyber-threat

probability. The other term in $P_{stop}(S_i)$, which is a linear combination of state variables F_i and C_i^{max} is adopted from [5], where $P_{stop}(F_i)$ and $P_{stop}(C_i^{max})$ represents the probability of cascade stop with F_i failures and capacity C_i^{max} respectively. The formulations were developed based on extensive cascading failure simulations over the IEEE 118 bus grid. Note that $P_{stop}(F_i)$ has a bowl-shape pattern defining three phases of a cascading failure and $P_{stop}(C_i)$ has an exponentially decreasing pattern that is high (low) when the maximum capacity of the failed transmission line is low (high).

$$P_{stop}(F_i) = \begin{cases} a_1 \left(\frac{a_2 M - F_i}{a_2 M} \right) + \epsilon & \text{if } 1 \leq F_i \leq a_2 M, \\ \epsilon & \text{if } a_2 M < F_i \leq 0.6M, \\ Q(F_i) & \text{if } 0.6M < F_i \leq M, \end{cases} \quad (4.17)$$

$$P_{stop}(C_i) = \max \left\{ a_3 \left(\frac{\max\{\mathcal{C}\} - C_i^{max}}{\max\{\mathcal{C}\}} \right)^4, a_4 \right\}. \quad (4.18)$$

The parameters $a_1, a_2, a_3, a_4, \epsilon$ defines the shape of the cascade stop probability. Parametric relationships between the parameters and the operating parameters of the grid were established in [1]. In both [5, 1], a fixed load-shedding constraint parameter, $\theta \in [0,1]$ was considered, which is the ratio of loads where load shedding is restricted and the total load of the grid. Intuitively, the constraint of load shedding is negatively correlated with the level of inter-connectivity, i.e., increasing the level of interconnectivity increases the availability of information, which in turn decreases the constraint on load shedding. With power node failures, the corresponding interconnection is also lost, which increases the constraint on implemented load-shedding in the grid. To capture this dynamical behavior, we differ from fixed θ in [5] and define θ as

In the proposed model, we consider load shedding as a dynamic parameter that

depends on three components: the level of inter-connectivity, system operators' error probability, and the number of line failures. We define load shedding, θ for any state as follows:

$$\theta = \min\left(1, \text{HEP}(F_i, C_i^{max}, \Psi_i) + \left(1 - k + \frac{k l F}{N_{pc}^{max}}\right)\right), \quad (4.19)$$

The parameters a_1, a_2, ϵ and a_3, a_4 are kept as same as defined in (4.6) - (4.8) and (4.10) - (4.11) respectively.

In equation 4.19, $l = N/M$, where N is the number of power nodes (buses) and M is the number of transmission lines. Transition probabilities, $P(S_j|S_i)$ of the Markov chain, are calculated as follows:

$$P(S_j|S_i) = \begin{cases} P_{cont}(S_i)(1 - P_{hc}(S_i)) \frac{u(\Psi_j)}{\sum_{l:\Psi_l \geq \Psi_i} u(\Psi_l)} \\ \quad \text{if } C_j^{max} = C_i^{max}, \\ P_{cont}(S_i)P_{hc}(S_i) \frac{w(C_j^{max})}{\sum_{m:C_m > C_i^{max}} w(C_m)} \frac{u(\Psi_j)}{\sum_{l:\Psi_l \geq \Psi_i} u(\Psi_l)} \\ \quad \text{if } C_j^{max} > C_i^{max}, \end{cases} \quad (4.20)$$

where $P_{cont}(S_i) = 1 - P_{stop}(S_i)$ and

$$P_{hc}(S_i) = \min(1, \alpha(F_i + \beta)^3), \quad (4.21)$$

where α and β are constants. Moreover, $P_{hc}(S_i)$ is the transition probability to a state with higher capacity and the $w(C_k)$'s are capacity weight parameters calculated from power grid simulations in [5]. We adopt the formulation of the human-error probability from [88] and ψ_i using (3), and after each failure in the power grid, we update the parameter θ_i parameter using 4.19.

Now, for any given initial condition, we can calculate the limiting distribution

of the failed transmission lines for a given initial state. Let, $\boldsymbol{\pi}_0$ be a vector that denotes the initial state S_0 of the Markov chain. Next, let the vector $\boldsymbol{\pi}^{(S_0)} = (\pi_i^{(S_0)}, i = 1, \dots, 2M|\mathcal{C}||\Psi|)$ represents the limiting distribution of the Markov chain starting from the state S_0 , where $\pi_i^{(S_0)}$ is the steady-state probability of the state S_i . Hence, $\boldsymbol{\pi}^{(S_0)} = \boldsymbol{\pi}_0 \lim_{k \rightarrow \infty} \mathbb{P}^k$. The conditional probability that a power grid eventually reaches a steady-state with F_i failures from an initial state S_0 is defined as

$$p(F_i|S_0) = \sum_{n=1}^{|\mathcal{C}||\Psi|} \pi_{2(F_i-1)|\mathcal{C}||\Psi|+2n}^{(S_0)}. \quad (4.22)$$

For a more detailed understanding regarding cumulative transmission-line failure probability and blackout size, we refer the reader to [5]. Using the distribution of blackout sizes, we can calculate the expected number of transmission-line failures, $\mathbb{E}[F_i|S_0]$, given the initial condition, S_0 as follows,

$$\mathbb{E}[F_i|S_0] = \sum_{F_i=1}^M F_i p(F_i|S_0). \quad (4.23)$$

4.4 Results

In this section, we share the capabilities of the I-SASE model using the IEEE 188-bus topology used in [5]. Since the SASE [5] and hSASE [1] models are already validated using power grid simulations over IEEE 118 and IEEE300 bus test cases and published in IEEE transactions on power systems previously, in this work we do not show additional validation to save space. For validation of the original SASE model, interested readers are suggested to read [5]. However, a comparison between the proposed model with a SASE and hSASE model is shown here to share the dissimilarities between observed outcomes along with additional capabilities.

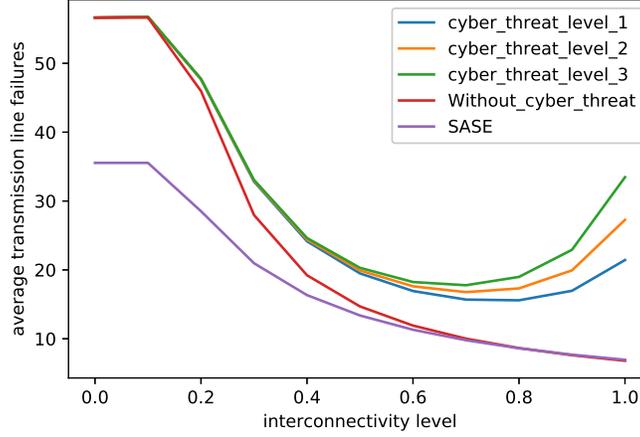


Figure 4.12: Expected transmission line failures for various inter-connectivity, and cyber threat level. We considered $F_i=2$, $C^{max}=20$, $r =0.7$, $e=0.1$ and nominal human error level. We can observe a point of diminishing return when cyber threat is considered.

4.4.1 Optimal power-communication inter-connectivity

One benefit of the proposed I-SASE model compared to the existing literature is that it finds an optimal level of inter-connectivity between the power grid and the associated communication network given human error probability and level of cyber threat. We use the expected value of transmission line failures as a measure of grid reliability, which can be calculated using equation (14). Increasing interconnectivity reduces constraint on load shedding but increases cyber threat probability. Thus, Due to the convex nature of the problem, there exists an optimal inter-connectivity that minimizes $E[F_i|S_0]$ subject to $0 \leq \theta, \Psi, HEP \leq 1$ and for a given grid condition during cascade initiation. Again, increasing the number of transmission line failures reduces inter-connectivity, which in turn, increases load shedding constraint, system operators' error probability but reduces cyber threat. Initially, increasing the level of inter-connectivity reduces the constraint on load shedding, which reduces the expected value of transmission line failures. As we increase inter-connectivity, cyber threat starts to play its role, and there

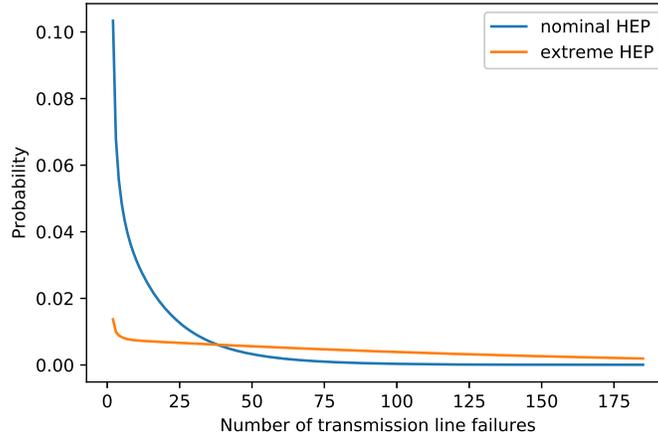


Figure 4.13: Distribution of the blackout size given initial condition of the grid for various nominal and extreme system operator error scenario,. We considered $F_i=2$, $C_i^{max}=20$, $\Psi=1$, $r =0.7$, $e=0.1$

exists a point of diminishing return beyond which the cyber threat dominates over capability to implement load shedding. We plot the $E[F_i|S_0]$ against the various level of inter-connectivity for different cyber threat level in Figure 5. Here we also plotted $E[F_i|S_0]$ when the cyber-threat probability is zero, i.e., without considering any cyber threat (marked by the red line). Notice that when cyber threat is not considered, increasing inter-connectivity decreases $E[F_i|S_0]$, and the optimal value of inter-connectivity is one, which matches the observed behavior in [7]. When cyber threat and nominal system operator error probability is considered, for the chosen initial condition ($F_i=2$, $C_i^{max}=20$, $r =0.7$, $e=0.1$), we observed optimal inter-connectivity at 80%.

4.4.2 Comparison with similar models

Recall that our proposed model is a generalization of the SASE model in [5]. Thus, if we exclude the effect of cyber threat and system operator errors, the model collapses to the SASE model. In that case, θ is a fixed parameter and does not change with the dynamics of the power grid. We have plotted $E[F_i|S_0]$ calculated using SASE model against the various level of inter-connectivity in Figure 4.12.

Note that when SASE model is considered (without cyber threat and operators' error), inter-connectivity level and load shedding constraint parameters are fully correlated. We can observe in Figure 4.12 that when inter-connectivity level is high, the line plots for SASE (purple) and without cyber threat (red) is a close match but not for when inter-connectivity level is low. The reason for this is that, for without cyber threat (red) case, system operators' error is considered, which created the difference between the two plots. However, when inter-connectivity level is high, load shedding capability increases significantly which supersedes the effect of operator error.

4.4.3 System operators' role on cascading failure

Recall that in Figure 4.10, we plotted operators' error probability against state indexes, which is monotone in nature. Error in operators' decision-taking adds more constraint on load shedding capability, which reduces the probability of cascade stopping and hence increases $E[F_i|S_0]$. The effect of various operator performance scenarios can be visualized in Figure 4.13. We observe that an exponential distribution for nominal HEP during a heavy tail distribution of extreme HEP.

4.5 Summary of this chapter

In this chapter, a cyber threat and system operator error aware Markov chain based model is presented for analytically predicting cascading failures in the power grid. Such a model is a significant enhancement and more realistic compared to the existing models for analyzing cascading failures probabilistically. During the propagation of cascading failures, the capability of the system operators, and the reliability of the available information can impact the dynamics of cascading failures significantly, which has been captured effectively in this model. Both benefits and harm of interdependency between power grid layers were captured through the dynamic load-shedding parameter and the cyber threat variable. Sys-

tem operators' error in taking the right actions at the right time was captured through load shedding constraints. Using this I-SASE model, statistics such as the distribution of transmission line failures conditional on initial condition in the steady-state, Expected values of the transmission line failures can be predicted, which can be useful to utilities for designing the grid. Additionally, an optimal power-communication inter-connectivity level given initial conditions of the grid can be calculated, which also can be a useful measure of reliability for the grid. One drawback of the model is that due to the unavailability of full grid network simulation data, the analytical model for cyber threat transitions could not be calculated analytically. Future efforts in validating the model in a live testbed would be crucial before implementing the model in a live grid network.

Chapter 5

Dynamics of Transmission Capacity and Load Loss during Cascading Failures in Power Grids

In this chapter, an analytical model is proposed to predict the average transmission-capacity loss and load loss during a cascading failure as a function of time and their steady-state values. Cascading failures in the power grid are described using a Markov-chain approach, in which the state transition probabilities depend on the number and capacities of the failed lines. The transition matrix is characterized parametrically using Monte Carlo simulations of cascading failures in the power grid. The severity of cascading failure is estimated using two metrics: the expected number of transmission-line failures and the amount of load shedding/load loss (inferred from the average transmission capacity loss) in the steady-state. These two metrics provide critical information regarding the severity of a cascading failure in a power grid (in terms of both the distribution of blackout sizes and the amounts of load shedding). One of the benefits of this model is that it enables the understanding of the effect of initial failures and the operating parameters of the power grid on cascading failures.

5.1 Introduction

In this chapter, we propose a model based on a Markov-chain to characterize the dynamics of cascading failures and calculate the severity of a cascade using two widely used metrics, i.e., the number of failed transmission-lines and the amount of load shedding/load loss (calculated from the average transmission capacity loss) during cascading failures.

The contribution of this work is three-fold. First, we construct the transition

matrix of the reduced state space using a realistic setting derived from power flow simulations using MATPOWER [79]. While in the SASE model [5], the allowed transitions in the next state were either to the same capacity or to a higher capacity of the transmission lines, here we consider transitions to any capacity in the next state. As a result, the calculation of the steady-state probabilities is more detailed compared to the SASE model. Second, our approach of calculating the transition matrix allows us to calculate the average transmission capacity loss (ATCL) by using a recursion technique. Third, we use numerical simulations to show that there is a linear correlation between the amount of load shedding and the cumulative capacity of the failed transmission lines during cascading failures. Hence, by using linear regression, we can infer the amount of load shed from an analytical calculation of ATCL. Most probabilistic models consider the distribution of the transmission-line failures as a metric of severity [13]. Using our model, we obtain the distribution of the blackout size and calculate the amount of load loss, which has been identified as the two critical metrics for evaluating the severity of a cascade in [58]. We name this model cSASE (capacity-SASE) to emphasize its predictive capability of tracking ATCL loss during a cascading failure.

5.2 cSASE model

5.2.1 Review of the simulation framework

The cSASE model is driven by numerous cascading failure simulations performed on the IEEE 118-bus system to understand cascading failure behavior. We use MATPOWER, a package of MATLAB, which provides a solution to the steady-state DC power-flow optimization problem. MATPOWER [79] has been used in several previous works [5, 1, 32] to analyze cascading failures in power grids. Most of the state-of-the-art probabilistic models on cascading failures use the DC power flow model for its simplicity yet effectiveness [29, 7].

In what follows, we consider a failure of a transmission line occurs when the power flow through the transmission line exceeds the maximum allowable power-flow limit of that transmission line. This maximum allowable power-flow limit is the capacity of that particular transmission line. Once the flow on a transmission line exceeds the capacity, we fail that line and re-calculate the power-flow again using the remaining transmission lines and repeat this process until we find no overloaded transmission lines. In our simulations, we consider one transmission-line failure occurs at each unit time. If multiple transmission lines exceed the capacity threshold, we fail the line with the maximum deviation from the overflow threshold. Since the power grid needs to balance generation and load, the failure of the overloaded transmission lines can lead to a cascade of failures in the successive time steps. To provide further context, similar to [5], we consider three power-grid operating parameters. Namely, we define the power grid loading level, $r \in [0,1]$, as the ratio of the total load demand and the total generation capacity. The load-shedding constraint $\theta \in [0, 1]$ is defined as the ratio between the total uncontrollable load (loads that do not participate in load shedding) and the total load in the power grid. The capacity estimation error, $e \in [0, 0.5]$, quantifies the error by the control center in estimating the actual capacity of the lines. Note that simulation results show that transmission-line failures increase with the r , e , and θ parameters. Further, in our simulations, in addition to variations and randomness in the initial failure, hidden failures are modeled by adding a small probability of failure for the lines adjacent to failed lines. From the simulations, we observe that depending on the grid topology, power grid operating parameters, and initial disturbances, the severity of the cascading failure varies from no cascading failures to complete power grid blackouts (a blackout occurs when a significant portion of the grid is failed).

5.2.2 State variables of the Markov chain

The state of the cSASE model is defined as $S_i = (F_i, C_i, I_i)$, where F_i, C_i, I_i are the state variables of the Markov chain. Here, F_i is the number of transmission-line failures at state S_i . To stay consistent with the simulation, we allow one transmission-line failure in our Markov chain at a time, which corresponds to setting a time step as the time needed for only one line failure to occur. C_i represents the capacity of the latest failed transmission line. Based on the simulation, we group the IEEE 118-bus system into 5 possible power-flow capacities; namely, $\mathcal{C} = \{20\text{MW}, 60\text{MW}, 120\text{MW}, 200\text{MW}, 332\text{MW}\}$ representing the quantized power-flow capacities based on transmission line) associated with the 186 transmission lines of the IEEE 118-bus topology. For a different topology, the set of power flow capacities may be different in size and contain different values. When a new transmission line fails at state S_j , the capacity of that line would be $C_j \in \mathcal{C}$. To capture this phenomenon, we consider every possible one-step transition in the transition matrix from the previous state. Hence the main difference with the SASE model is that we keep track of the capacities of the lines that fail at any time, instead of the maximum capacity of the failed lines up to that point.

5.2.3 Formulating the transition matrix

The transition matrix of the Markov chain is a $2M|\mathcal{C}| \times 2M|\mathcal{C}|$ matrix, where M is the total number of transmission lines and $|\mathcal{C}|$ is the cardinality of the set of capacities (five in this work) and 2 accounts for the binary variable I . In our model, we do not allow failed lines to recover. So, the number of transmission-line failure, F_i , increases monotonically at a rate of one failure per unit time. Since, F_i increases monotonically with no healing capabilities, the transition matrix \mathbb{P} is a row-stochastic upper diagonal matrix. The state transition probability, $f(S_j|S_i)$,

from state $S_i = (F_i, C_i, I_i)$ to the next state $S_j = (F_j, C_j, I_j)$, is defined below

$$f(S_j|S_i) = \begin{cases} 1 & \text{if } F_j = F_i, C_j = C_i, I_j = I_i = 1, \\ P_{stop}(S_i) & \text{if } F_j = F_i, C_j = C_i, I_j = 1, \\ P(S_j|S_i) & \text{if } F_j = F_i + 1, C_j \in \mathcal{C}, I_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.1)$$

Here, $P_{stop}(S_i)$ is the probability of the cascading stopping at state S_i , and $P(S_j|S_i)$ is the probability of a new transmission-line failure with capacity C_j given the F_i transmission-line failures with latest capacity loss of C_i . We introduce the marginal probability, $P_{F_i}^{C_i}$, which is the probability of a transmission-line failure with capacity C_i when the number of failed line is F_i . Based on our set of power flow capacities \mathcal{C} and power grid simulations with 5000 random simulations over the IEEE 118-bus system, we obtain the following probabilities of transmission-line failures: $P_1^{20} = 0.898$, $P_1^{60} = 0.076$, $P_1^{120} = 0.020$, $P_1^{200} = 0.005$, $P_1^{332} = 0.001$. We use this set of marginal probabilities as the initial input to our model. Consider an initial event with one or more transmission-line failures, i.e., with a known F_i and C_i . A cascading failure stops when the Markov chain reaches an absorbing state (no additional failures), i.e., for a state with $P_{stop}(S_i) = 1$. Note that, $P_{stop}(S_i)$ is a key parameter of the Markov chain model and the characterization of $P_{stop}(S_i)$ was discussed extensively in [5].

Here, $P_{stop}(S_i)$ is defined as a linear combination of $P_{stop}(F_i)$ (the probability of cascade-stop with F_i failures) and $P_{stop}(C_i)$ (the probability of cascade-stop with capacity C_i). We adopt the formulation of $P_{stop}(F_i)$ from [5]. However, power-grid simulations show that $P_{stop}(C_i)$ is high (low) when the capacity of the failed transmission line is low (high). Hence, $P_{stop}(C_i)$ is defined as

Table 5.1: α, β, γ for different values of F_i, r, e, θ

F_i	r	e	θ	α	β	γ
2	0.7	0.4	0.1	44.33	-1.37	-1.19e-05
2	0.85	0.45	0.2	4159.02	-2.84	-1.95e-05
3	0.6	0.3	0.2	2919.47	-2.69	-1.08e-05
3	0.85	0.45	0.2	2167.98	-2.63	-2.32e-05
4	0.85	0.45	0.2	2048.77	-2.60	-1.44e-05
5	0.85	0.45	0.2	1620.52	-2.51	-7.2e-06

$$P_{stop}(C_i) = \max \left\{ a_3 \left(\frac{C_i - \max\{\mathcal{C}\}}{\max\{\mathcal{C}\}} \right)^4, a_4 \right\}. \quad (5.2)$$

Here a_3 and a_4 are scaling parameters for $P_{stop}(C_i)$ in addition to a_1 and a_2 for $P_{stop}(F_i)$, which are modeled parametrically using power grid simulations [5].

We formulate $P(S_j|S_i)$ using $P_{stop}(S_i)$ (defined above) as

$$P(S_j|S_i) = \left(\alpha(C_j)^\beta (1 + \gamma C_i(C_j - C_{th})) \right) (1 - P_{stop}(S_i)). \quad (5.3)$$

The parameters α, β, γ and C_{th} are calculated as done in [89]. Here, C_{th} is a threshold in capacity values, where transmission lines with capacity lower than C_{th} are more vulnerable to failure in the next step. For the chosen set of capacities, a value of $C_{th} = 41\text{MW}$ was observed from power grid simulations similar to [89]. For reproducibility, we list the values of α, β, γ calculated from the fitted power grid simulation data for different values of r, e, θ and initial F_i in Table 1. This completes the calculation of $f(S_j|S_i)$ in (1) for all S_i, S_j .

5.2.4 Calculating the average transmission capacity loss in the steady state

ATCL is a piece of critical information for measuring the severity of the cascading failure and regarding the reliability of the power grid. NERC defines a large cascading failure that occurs in the power grid when the total load loss exceeds

300MW[32]. Next, we define our approach of calculating the ATCL, which will allow us to calculate the load loss during cascading failure.

We introduce the following recursion to calculate the average transmission capacity loss during cascading failures in the power grid as

$$ATC_{F_j} = ATC_{F_i} + ACL_{F_j}, \quad (5.4)$$

where ACL_{F_i} is the ATCL in the current state with F_i failures, and ATC_{F_j} is the average total capacity (ATC) loss with a total of F_j failures during cascading failures. To calculate ACL_{F_j} , we need the marginal probabilities of initial line-failures with capacity C_i at the current state. After the occurrence of an initial event, we calculate the marginal probability at successive steps as follows,

$$P_{F_j}^{C_j} = \sum_{C_i \in \mathcal{C}} P(C_j | C_i) P(C_i), \quad (5.5)$$

where $P(C_j | C_i) = P(S_j | S_i)$ is obtained using (5). Here, $P(C_j | C_i)$ equals $P(S_j | S_i)$ because of our definition of the transition matrix in (1). Note that, $P(C_j | C_i)$ depends on the particular bowl shape of $P_{stop}(S_i)$ and varies with F_i and C_i . Then, we calculate ACL_{F_j} as follows

$$ACL_{F_j} = \sum_{C_j \in \mathcal{C}} C_j P_{F_j}^{C_j}. \quad (5.6)$$

Now, for any given initial condition, we can calculate the distribution of the failed transmission lines for various operating conditions (r, e, θ) of the power grid. Let, $\boldsymbol{\pi}_0$ be a vector that denotes the initial state S_0 at time $k = 0$. Then $\boldsymbol{\pi} = \boldsymbol{\pi}_i, (i = 1, \dots, 2M|\mathcal{C}|)$ represents the limiting distribution, i.e., $\boldsymbol{\pi}_i = \boldsymbol{\pi}_0 \lim_{k \rightarrow \infty} \mathbb{P}^k$. The conditional probability that a power grid eventually reaches a state with F_i

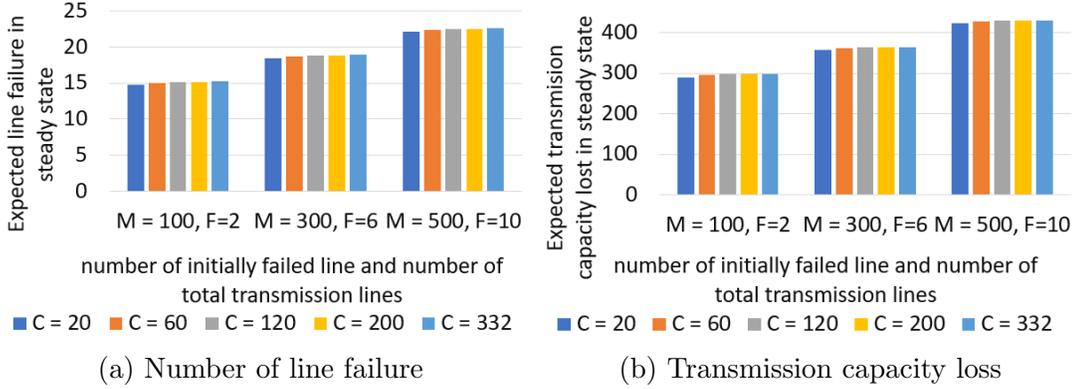


Figure 5.1: Expected number of line failure and expected transmission capacity loss for different total number of transmission lines M and initially failed transmission lines F . Here, C_i is the capacity of the latest transmission line failed among F transmission lines and $r = 0.85$, $e = 0.45$, $\theta = 0.2$.

failures from an initial state S_0 , is defined as

$$p(F_i|S_0) = \sum_{i=1}^{|C|} \pi_{2(F_i-1)|C|+2i}. \quad (5.7)$$

For a more detailed understanding regarding cumulative transmission-line failure probability and blackout size, we refer the reader to [5]. We then calculate the expected number of transmission-line failures, $\mathbb{E}[F_i|S_0]$, given the initial condition, S_0 as follows,

$$\mathbb{E}[F_i|S_0] = \sum_{F_i=1}^M F_i p(F_i|S_0). \quad (5.8)$$

Finally, we can calculate the expected total capacity loss for any given initial condition, S_0 , using

$$\mathbb{E}[ATC_{F_i}|S_0] = \sum_{F_i=1}^M ATC_{F_i} p(F_i|S_0). \quad (5.9)$$

5.3 Results

5.3.1 Expected number of transmission-line failures and average transmission capacity loss

In Figures 5.1(a) and 5.1(b), we show the trend of the expected number of line failures and ATCL during a cascading failure for various initial conditions. One great advantage of the SASE model is its scalability, and we can also scale the cSASE model for any grid topology. We consider fixed $\alpha, \beta, \gamma, r, e,$ and θ for the snapshots in Figures. 5.1(a) and 5.1(b). We increase the initial number of failed transmission lines proportionally as we increase M . Figures 5.1(a) and 5.1(b) show the expected line failure and ATCL increase with an increase in $F, C_i,$ and M . Similarly, we calculate the expected line failures using the SASE model. The expected number of line failures in the SASE model is significantly high compared to the cSASE model (for $F = 4$ and $M = 200$, expected line failure is 26 in the SASE compared to 17 in the cSASE). The reason is that in the SASE model, a transition to line failures with lower capacity in the Markov chain state space is not allowed. Once a high-capacity transmission-line fails, there is no transition to a low-capacity transmission line failure in the next step. In the cSASE model, we allow all possible one-step transitions. Namely, even if there is a high capacity transmission-line failure at the current state, a low capacity line may fail at the next step. This difference causes the $P_{stop}(S_i)$'s to be higher in the cSASE compared to the SASE model.

5.3.2 Predicting load loss from average transmission capacity loss

We perform extensive cascading failure simulations on the IEEE 118-bus system using various power grid conditions. We randomly select the value of r from the vector $\{0.5, 0.6, 0.7, 0.8, 0.9\}$, e randomly from the vector $\{0.05, 0.1, 0.15,$

0.2, 0.25}, θ randomly from the vector {0.05, 0.1, 0.15, 0.2, 0.25}, and set $F_i=2$ for each iteration of the simulation. The purpose of taking vectors for different operating parameters is to generate different combinations of initial operating conditions randomly with two transmission line failures. Note that, in simulation, we use the DC optimal power flow algorithm, which is capable of implementing load shedding based on the cost of generation and load. In the simulation, we set the cost of load shedding ten times higher than the cost of generation. Similar to [5], we use dispatchable loads in the MATPOWER solver. To do that, we add additional buses to separate buses with both load and generation in the IEEE 118-bus system and connect the new buses with the added negative generators that result from the dispatchable load (two or more generators cannot be added in the same bus in MATPOWER solver). As described earlier, we fail one transmission line in each step of the cascade and redistribute the power flow. At the same time, the MATPOWER solver implements load-shedding if required. We collect measurement data containing cumulative transmission capacity of the failed lines and the cumulative amount of load shed (load loss) from 76280 cascading failures simulations from MATPOWER. Then, we calculate the average load shed for a various cumulative capacity of the failed lines, which is plotted in Figure. 5.4. Observe that the amount of load shed is linearly correlated with the cumulative capacity of the failed transmission lines. The red line in Figure 5.2 shows the linear trend line calculated using linear regression with the slope coefficient and the y-intercept being 0.1378 and - 33.461, respectively, for our data set. We see from Figure 5.2 that the load shed variance is higher for the higher values of the cumulative capacity of the failed transmission lines due to low sample sizes for each bin. The linear relationship between the amount of load shed and cumulative capacity of the failed lines allows for a calculation of the amount of load shed from the cumulative capacity of the failed transmission lines analytically. This observation

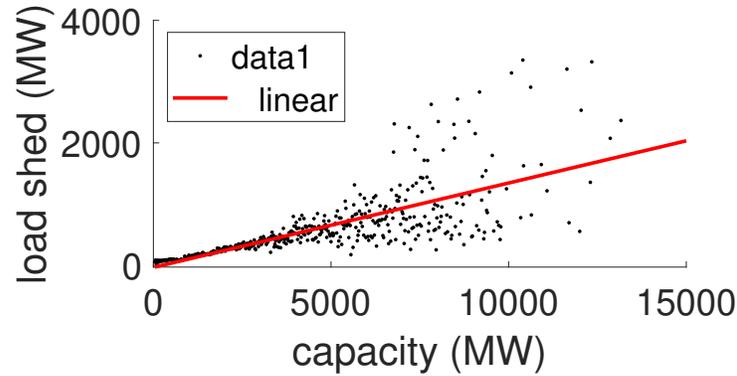


Figure 5.2: Predicting average load shed from cumulative transmission capacity of the failed transmission lines.

	M =100	M =200	M =300	M =400	M =500
F=2	290	262	236	214	196
F=3	315	296	270	247	228
F=4	336	324	301	279	259
F=5	355	349	330	309	289
F=6	375	371	357	337	318

Figure 5.3: Expected transmission capacity loss during cascading failures for different M , F (red indicates a cascade). Here $C_i = 20MW$, $r = 0.85$, $e = 0.45$, and $\theta = 0.2$.

is critical since the Markov chain simulation, and we obtain the ATCL for a given initial condition. This allows us to use (i.e., the amount of load shed = $0.1378 \times$ ATCL - 33.461) linear regression to get the amount of load shed from the ATCL.

5.3.3 Critical initial conditions for various grid sizes

We show the ATCL for various initial conditions ($C_i = 20MW$, $r = 0.85$, $e = 0.45$ and $\theta = 0.2$) and number of transmission lines (F) in Figure 5.3. We use a threshold of 300MW for ATCL (which corresponds to a small amount of load loss), above which we consider a cascading failure event. Note that the ATCL value was chosen arbitrarily to classify cascade events. The table in Figure 5.3 shows the severity of cascading failure for different values of M and F (e.g., the red-colored zones indicate a cascade). Observe that, when the total number of

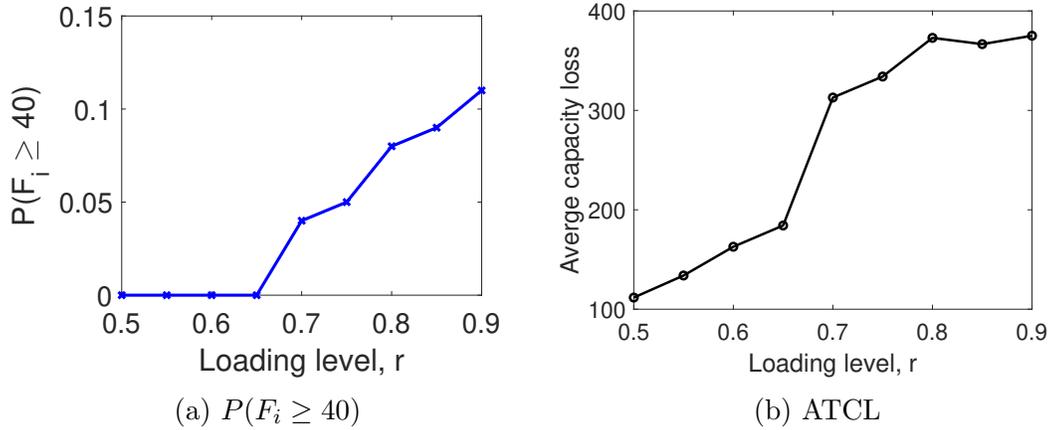


Figure 5.4: $P(F_i \geq 40)$ and ATCL for various r , $e = 0.45$, $\theta = 0.2$. For $r \geq 0.65$, probability of a cascade increases with r .

transmission lines in the power grid M is 100, ATCL is less than 300MW only for initial $F = 2$. Hence, when $M \leq 100$, an initial failure with more than two transmission-lines will lead to a cascading failure on average. On the other hand, when $M = 500$, power-grids are resilient to cascading failures even though there are five initial failures. Note that the results in Figure 5.3 are dependent on the topology of the power grid.

In Figure 5.4a, we show show the cumulative probability of 40 or more transmission-line failures, $P(F_i \geq 40)$ for various r and in Figure 5.4b the average transmission capacity loss for various r . Here the parameters $a_1, a_2, a_3, a_4, \alpha, \beta$, and γ are approximated from power grid simulation data and curve fitting. Figure 5.4 shows that the cumulative probability of 40 or more line failures increases sharply when $r > 0.65$. ATCL also jumps from 184MW ($r = 0.65$) to 313MW ($r = 0.7$). This indicates the critical load level for this particular setting. This critical load level is critical in the cascading-failure analysis because beyond this level, failures follow a power-law distribution rather than an exponential distribution. Thus the probability of having a large cascade is higher beyond this critical load level. Note that this critical load level varies with the power grid topology as well as the values of $a_1, a_2, a_3, a_4, \alpha, \beta$, and γ , i.e., $P_{stop}(S_i)$ and $P(S_j|S_i)$.

Chapter 6

Damage of initial assets can have different consequences

Cascading failures in the power grid are heavily dependent upon the initial stressor event that induces failures in the power grid and initiates a chain of events. A stressor event can be a natural disaster or human-made sabotage attack or error. Power grid parameters such as the number of failed transmission lines, the capacity of the failed transmission lines, the loading level in the power grid, the ability to implement load-shedding, collectively affect the cascading behavior following an initial event. Moreover, the geographical correlation among failures during an initial event can amplify cascading failures [8, 90, 34]. A combination of parameters determines the initial failures of the power grid due to the occurrence of an initial stressor event and can lead to blackouts of various sizes in the power grid. Therefore, to model the dynamics of the cascading failure in the power grid, it is essential to investigate the impact of power grid parameter responses upon an initial stressor event. In this chapter, we study the influence of the initial conditions that conduce the cascading failures in the power grid. We formulate the impact of stressor(s) events analytically using Gaussian, circular, and linear degradation functions, which result in initial failures in the power grid. We perform simulations on the IEEE 118-bus and IEEE 300-bus topology using a power flow simulator to observe the impact of initial stressor(s) event and power grid parameters on cascading failures in the power grid. Simulations show that there is a linear relationship between initial failures in the power grid and stressor(s) intensity. Then, we observe the impact of the number of initially failed transmission lines and the total capacity of the initially failed transmission lines on cascading failures in the power grid. Moreover, we use the power flow simulator

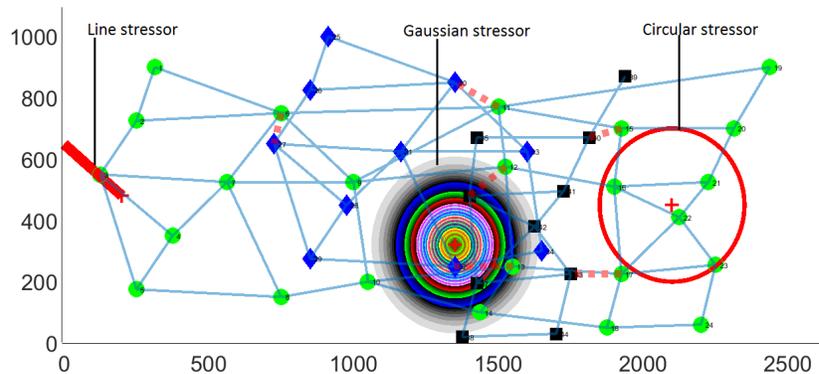


Figure 6.1: Gaussian, circular and linear attacks mapped into a 2-D topological space. Failures in the power grid depends on the intensity of the stressor(s). Initial transmission line failures in the power grid calculated after an attack is simulated [8].

to investigate the impact of power grid loading level and load-shedding during the initial event. Our simulation results show that a combination of power grid parameters influence cascading failures drastically. These parameters include the number of failed transmission lines, the total capacity of the failed transmission lines, number of geographical stressor(s) locations, failed transmission lines in each stressor location, the intensity of the stressor(s), the power grid loading level, the load-shedding constraints. Increasing the values of these parameters during an initial event increases the probability of cascading, i.e., increases the probability of blackout-size in the power grid. Simulation results suggest that the initial condition of the power grid during a stressor(s) event is very crucial; hence, this work paves a way to study and minimize the impact of cascading failures with carefully designing the grid considering these effects.

6.1 Modeling the initial failures due to the stressor(s) and impact of the stressor(s) on cascading failures in power grid

In recent years, researchers contributed significantly to model the cascading failures in the power grid. To the best of our knowledge, most of the works done on the probabilistic modeling of cascading failures consider arbitrary initial failures

and then focus on modeling the propagation of failures. However, fewer efforts are made to observe the impact of various initial conditions that lead to cascading failures, which is the crucial contribution of this chapter. We map the intensity of stressor(s) events with failures in the power grid. No notable extensive analysis has been done to show the correlation between the status of power-grid parameters during an initial stressor(s) event and failures in the power grid that leads to cascading failures. Our work can map the correlation between an initial stressor event and cascading failures in the power grid; thus, this work can investigate the cascading failure behavior of the power grid more realistically compared to other works. In this section, we map the initial transmission line failures in the power grid with stressor intensities.

6.1.1 Modeling the initial failures due to stressor(s)

Multiple stressors can occur in one geographical location, or they can spread over different geographical areas. These stressor(s) events can range from natural disasters (e.g., tornado, cyclone, earthquake) to intentional human-made attacks (e.g., use of weapons of mass destruction (WMDs), High altitude electromagnetic pulses (HEMPs), cyber-attack in the communication layer of the power grid. These events can lead to initial disturbances in the power grid, which may include the transmission line failures, generator loss, or failures in the communication system. These initial failures can act as a trigger for initiating cascading failures in the power grid. In this chapter, we have used spatially-homogeneous stressor(s) centers, which enables us to model multiple stressor(s) events at the same time. The spread of these stressors can vary depending on the intensity of the stressor(s). We use Gaussian, circular, and linear degradation functions, which can reasonably characterize various real-world stressor(s) [8]. The intensity of the Gaussian stressor degrades according to the Gaussian function as the spatial distance from the location of occurrence increases. The intensity of the function has

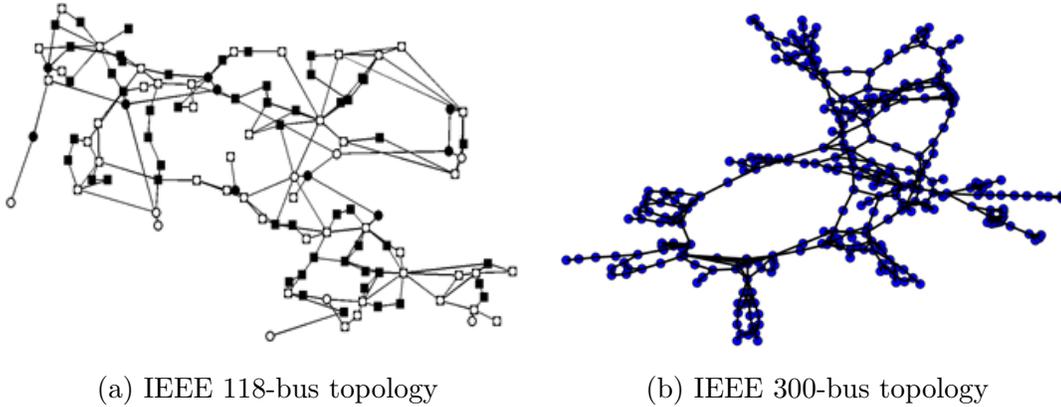


Figure 6.2: IEEE 118-bus and 300-bus topology

a peak at the mean of the degradation function. Two parameters entirely describe a circular degradation function: radius of the circle (r) and the intensity of the stressor at the center (I). The main difference between a Gaussian and a circular stressor is in their degradation function. For a Gaussian stressor, the intensity of the stressor degrades at e^{-d^2} , while for the circular stressor, it degrades with $1/d^2$. For the Gaussian case, d is the minimum distance from the stressor center to the point where intensity needs to be measured (e.g., Bus location, transmission line fault). Similarly, for the circular case, d is the distance from the stressor center to the point of intensity measure. Linear stressor(s) can be used to model natural disasters like tornadoes, which can occur in any geographical location with a shallow radius but having almost equal strength over the region it spreads. Figure 6.1 shows a realization of these three types of degradation functions over physical infrastructure. Attacks with the same intensity can lead to a different impact on the power grid (e.g., different transmission line failures) depending on the nature of the attack.

We denote the stressor(s) event by W and the stressor intensity at any point (x_i, y_i) from the center of the stressor(s) with $I_w(x_i, y_i) \geq 0$ (attack intensity is either zero or a positive number and cannot be negative). The shape of a stressor can be either Gaussian, circular, linear, or a combination of any of these over the

power grid topology. The stressor intensity degrades with distance from the center. To calculate the probability of line failures due to a stressor event, we divide each of the power grid transmission lines into N points (N can be infinity large, i.e., the distance between two adjacent points can be close to zero) and measure the stressor intensity at those points after the occurrence of a stressor event. We then take the maximum intensity calculated in those N points. We assume that if the maximum intensity at any point over the line crosses a certain threshold, then the line will fail. Here, we assume N to be sufficiently large. An alternative approach to calculating the maximum stressor intensity on a transmission line can be to calculate the minimum distance between the transmission line and the stressor center. Since the stressor intensity degrades over distance, it is intuitive that minimum distance from the stressor center would result in maximum intensity, with the peak intensity being at the center of the stressor(s). Hence, the maximum stressor intensity on a transmission line would be inversely proportional to the minimum distance between the transmission line and the stressor center. For a single stressor event occurred in a geographical location, we define the failure probability of a transmission line as:

$$p((B_i, B_j)|W = w) = \min\left(\max_{k \in \{1, \dots, N\}} I_w(x_k, y_k), 1\right), \quad (6.1)$$

where $p((B_i, B_j)|W = w)$ denotes the failure probability of a transmission line of the power grid, (B_i, B_j) is the transmission line from B_i th bus to B_j th bus, and (x_k, y_k) is the location of the k th point on (B_i, B_j) . For multiple stressor events occurring at the same time, the total stressor intensity at (x_k, y_k) is

$$p((B_i, B_j)|W = (w_1, \dots, w_L)) = \min\left(\left(\sum_{i=1}^L \max_{k \in \{1, \dots, N\}} I_{w_i}(x_k, y_k)\right), 1\right), \quad (6.2)$$

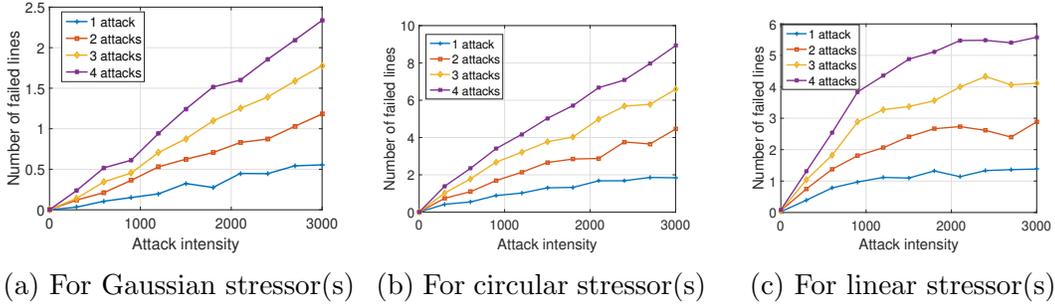


Figure 6.3: Average number of failed transmission lines in IEEE 118-bus topology due to Gaussian, circular and linear stressor(s) with various intensities.

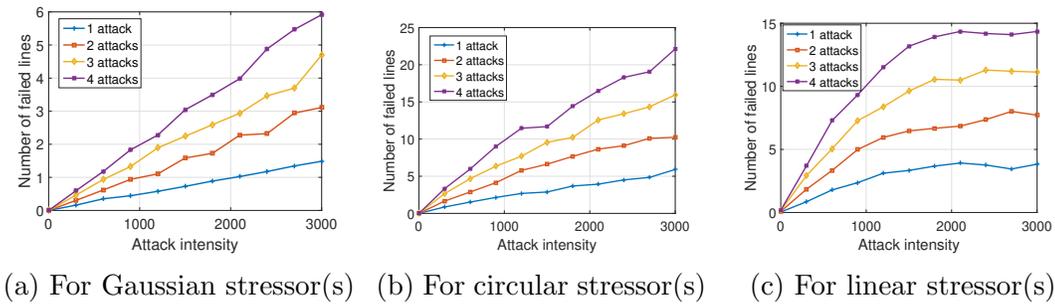


Figure 6.4: Average number of failed transmission lines in IEEE 300-bus topology due to Gaussian, circular and linear stressor(s) with various intensities.

where L denotes the number of stressors.

We calculate the total number of failed transmission lines in the power grid due to the occurrence of the stressor(s) using the measured individual transmission line probability. Similarly, we can calculate the bus (node) failure probability due to multiple stressor events using the following equation.

$$p((B_i)|W = (w_1, \dots, w_L)) = \min\left(\left(\sum_{i=1}^L I_{w_i}(x_k, y_k)\right), 1\right) \quad (6.3)$$

Now, considering the fact that initial failures of a network component does not depend on other components [90], the joint failure probability of the power grid transmission lines due to stressor(s) event can be represented using the product of their individual failure probabilities. Therefore, for a power grid with M

transmission lines we have

$$p\left(\left((B_1, B_2), \dots, (B_{M-1}, B_M)\right) | W = w\right) = \prod_{(B_i, B_j) \in V} p((B_i, B_j) | W = w), \quad (6.4)$$

where V is the collection of all transmission lines in the power grid. Depending on the geographical position and the intensity of the stressor(s), we obtain different initial transmission lines failures. Figure 6.3 and Figure 6.4 shows a plot of the average number of failed transmission lines due to the stressor(s) with Gaussian, circular, and linear degradation functions with various intensities. We obtain the average number of failed transmission lines using Monte-Carlo simulations over the IEEE 118-bus topology (186 transmission lines, Figure 6.2(a)) and IEEE 300-bus topology (411 transmission lines, Figure 6.2(b)) with 1000 sample realizations. In each sample realization, we generate stressor(s) at random locations (uniformly distributed) and calculate the intensity of stressor at every bus and transmission line using (6.1) and (6.2). Then we take the expectation of transmission line failures over the total realizations with a stressor(s) intensity for the three degradation functions. In both IEEE 118-bus and 300-bus cases, we can see that the expected number of failed transmission lines increases linearly with the increase in stressor(s) intensity. Again, it can be observed that for a particular stressor type and same attack intensity, for the IEEE 300-bus system, we get higher average failed lines compared to the IEEE 118-bus system. Aforementioned is because, for the IEEE 300-bus system, node density over the geographical region is higher compared to the IEEE 118-bus system. From Figures 6.3 and 6.4, it is visible that with the same stressor intensity, circular stressor creates the worst impact on both the IEEE 118-bus and IEEE 300-bus topology. On the contrary, Gaussian stressor has the least impact since Gaussian stressor(s) intensities decay at a faster rate (e^{-d^2}) compared to a circular stressor(s), which degrades with $1/d^2$ where $d < r$.

As shown above, the expectation of transmission line failure in the power grid has a linear relationship with stressor(s) intensities, i.e., the number of line failures increases linearly with the stressor(s) intensity for Gaussian, circular and linear degradation functions. With this relationship at hand, we now have a model that is capable of giving the initial failures in a power grid due to a stressor(s) event with various intensities. This is important because now we can predict the impact of a real-world natural disaster or human-made attacks. In the next section, we will use the obtained initial failures due to stressor(s) in optimal power flow simulator (MATPOWER) [79] and analyze cascading failures in the power grid.

6.1.2 Simulation framework

We use the MATPOWER [79] based CFS framework in our cascading failures analysis, which we described already in the previous chapters. It uses power-flow distribution framework and can give overloaded transmission lines, which was used in several previous works [84, 5, 1]. We consider a line failure when power flow through a line exceeds maximum allowable capacity through that line. Once we find an overflow in a transmission line, we fail that line and re-calculate optimal power flow (OPF) using the remaining transmission lines. In our simulations, we take one transmission line failure at a time. If multiple transmission lines exceed the capacity threshold, we fail the line with maximum capacity. We take 1000 random realizations and calculate associated transmission line failure probabilities due to the stressor(s) using (6.1) and (6.2). We use the same intensity of the stressor(s) for one turn of 1000 realizations and calculate the average number of failed transmission lines.

6.1.3 Impact of stressor(s) event on cascading failures

We use Gaussian, circular and linear stressor(s) over the IEEE 118-bus topology (Figure 6.2(a)), and consider these stressor(s) as initial events that may lead

to cascading failures in the power grid. We only show the results using the IEEE 118-bus topology here for space constraint. We perform Monte-Carlo simulation to analyze the impact of the stressor(s) on cascading failures in the power grid based on OPF analysis. Since transmission line failures increase linearly with stressor intensity, one stressor event can generate multiple transmission line failures if the stressor intensity is high. However, the line failures will exhibit clustering (failed lines will be close to each other). On the contrary, multiple stressors can initiate multiple failures, and the stressor locations can be distributed randomly (inhibition). Here, we define that a cascading failure event occurred if more than five percent additional transmission lines are failed after an initial stressor(s) event. If, $F_{threshold}$ is the threshold for a cascading event, $F_{initial}$ and M are the number of initially failed transmission lines and total number of transmission lines respectively then, $F_{threshold} = F_{initial} + 0.05M$. For a realization, if the total number of line failure exceeds $F_{threshold}$, we consider that as a cascading failure event. For example, if three transmission line fails due to a stressor(s) event, then we say a cascading failure event occurred if more than twelve transmission line fails for the IEEE 118-bus case, which has 186 transmission lines. From Figure 6.5, it is visible that inhibition of failures generates more cascading failure events than clustering, i.e., if the transmission line failures are randomly distributed, then there is a higher likelihood of cascading failures in the power grid. The reason for low cascading due to clustering is that the power grid has a better control mechanism to mitigate the impact of localized failures using load-shedding or islanding, as the location of the failed lines is very close to each other. Most of the probabilistic models consider random failures distributed over the power grid [5, 1, 7]. However, if the failed transmission lines are distributed (can be the result from multiple stressor events occurring at the same time in various locations), that in turn increases the probability of cascading failure in the power grid.

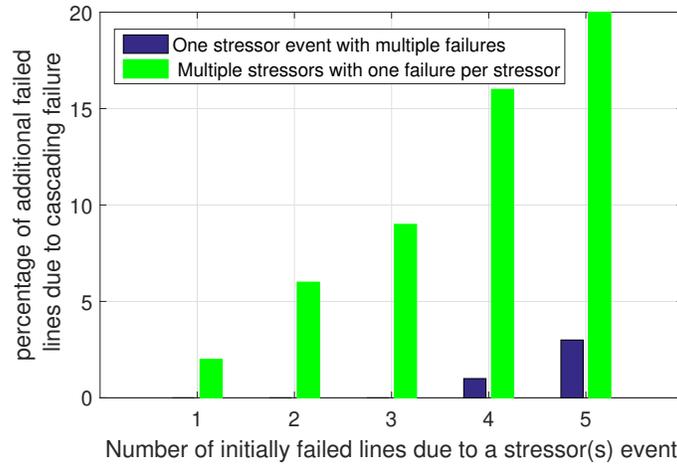


Figure 6.5: Number of failed transmission lines when one stressor location with multiple failures (blue) and considering randomly distributed failed transmission lines (green) where a stressor event contribute one transmission-line failure (we pick the line with maximum intensity to fail).

Figure 6.6 shows the simulation result for attacks with multiple transmission line failures. We can see that for the same number of transmission line failures, if we increase the number of attack points, the power grid becomes more cascade-prone than the previous case. Here, in Figure 6.6, we use linear curve fitting (blue, red, green, and orange lines represent various stressor(s)) to show the impact of inhibition clearly.

6.2 Impact of initial failures due to a stressor event

We now apply our initial failure model in MATPOWER OPF simulator to calculate the impact of stressor(s) events on cascading failures in the power grid. Simulations using the other IEEE topologies follow the same pattern.

6.2.1 Impact of number of failed transmission lines and capacity of the failed transmission lines

We define percentage of additional transmission lines lost due to the cascading failures as $\Delta_M / (M - M_{initial})$, where Δ_M = additional transmission lines lost due to cascading; M = total transmission lines of the power grid; $M_{initial}$ = number

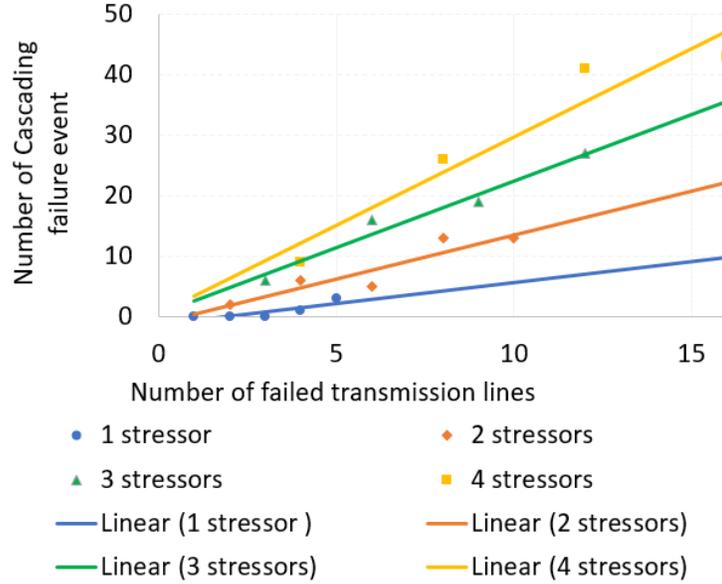


Figure 6.6: Number of cascading failure event in a power grid with different number of attack points and number of transmission line failures.

of transmission lines failed due to initial event. Similarly, percentage of additional capacity lost due to the cascading failures as $\Delta_C / (C_{total} - C_{initial})$, where $\Delta_C =$ additional capacity lost due to the cascading; $C_{total} =$ total capacity of the power grid; $C_{initial} =$ total capacity of the initially failed lines. Figure 6.7 represents the impact of various initially failed transmission lines of fixed total capacity and the total capacity of the failed transmission line during an initial event using OPF simulations. In Figure 6.7(a), we keep the total capacity of the failed lines as constant and then increase the number of failed transmission lines. We take randomly distributed line failures for 1000 samples in each case. These initial line failures are generated using random stressor events over the IEEE 118-bus topology. Our simulation results suggest that if the total capacity of the failed lines is fixed, an increase in the number of line failures makes the power grid more cascade-prone. In Figure 6.7(b), a similar type of simulation is done with a fixed number of failed transmission lines (randomly chosen from the 186 lines) while varying the total capacity of the failed lines. The results suggest that the

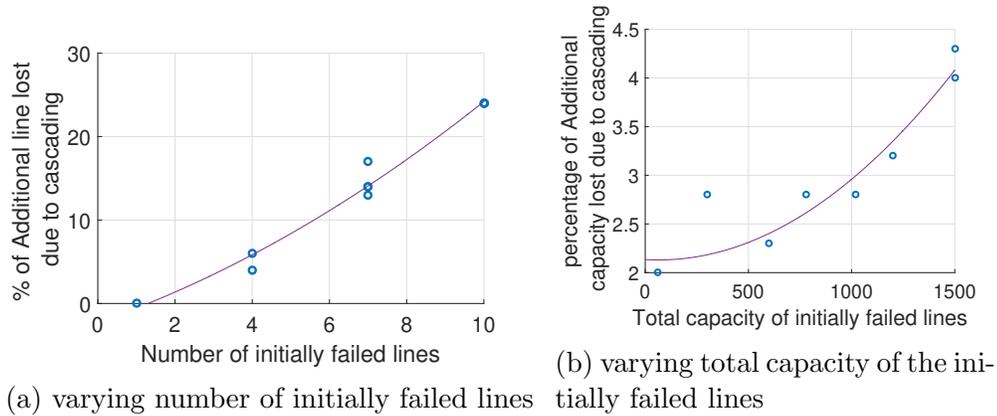


Figure 6.7: Relationship between the number of initially failed transmission line due to a stressor event with a percentage of additionally failed lines due to cascading when the total capacity of the failed transmission lines are fixed, and the total capacity of the initially failed transmission lines with additional capacity lost due to cascading when the number of the failed transmission lines are fixed.

percentage of additional capacity lost due to cascading failures increases if the total capacity of the initially failed lines is increased. Thus, we conclude that both numbers of initial line failures and the total capacity of the failed lines during a catastrophic event can lead to the cascading failures in the power grid.

6.2.2 Impact of power grid loading level and load-shedding constraint on cascading failures in power grids

Power-grid loading level, $l \in [0,1]$, is defined as the ratio of the total demand and the generation capacity of the power grid. The ratio of the uncontrollable loads (loads that do not participate in load shedding) and the total load in the power grid is termed the load-shedding constraint, denoted by $\theta \in [0,1]$. Here, the stress of the power grid increases as we increase l , and $\theta = 0$ implies no load shedding constraint while $\theta = 1$ indicates no load shedding can be implemented.

To observe the impact of l and θ , we consider a fixed number of initial transmission line failures in our simulation. We observe that when the power grid is highly stressed, it is more cascade-prone than when the grid is nominally stressed. Figure 6.8 shows a linear relationship between the average number of failed trans-

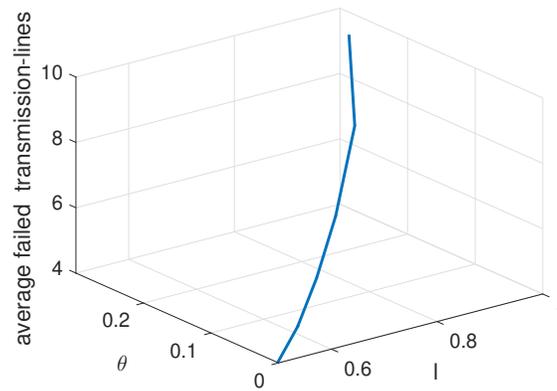


Figure 6.8: Dependence of the average number of failed transmission lines on power grid operating parameters (l and θ).

mission lines and the operating parameters. We can also observe that there is a critical operating point for both l and θ (approximately 0.8 and 0.2 for l and θ in our case). We observe a sharp increase in average cascading failures beyond this critical parameter setting. Similar observations were found in [5, 1].

Chapter 7

Predicting cascading failures in power grids using machine learning algorithms

Although there has been significant progress in modeling cascading failures in power grids, few works involved using machine learning algorithms. In this chapter, we classify cascading failures in the power grid that lead to large blackouts in power grids using machine learning algorithms. Since real-world cascading failure data is not available, we create a synthetic cascading failure simulator framework to generate cascading-failure data for various power grid operating parameters. We include the topological parameters such as edge betweenness centrality, the average shortest distance for various combinations of two transmission line failures in our dataset. Then we apply various machine learning algorithms to classify cascading failures and compare accuracy. Further, we use regressive models to predict the number of failed transmission line and the amount of load shedding. This data-driven technique is useful to quickly classify cascading failures based on the input power grid conditions. Hence, power grid design engineers can use this to increase the robustness of the grid.

In this chapter, we classify and predict cascading failure based on critical power grid attributes like power-flow capacity, edge betweenness centrality, demand loss, power grid loading, estimation errors, constraints on load-shedding. The contribution of this chapter is three-fold. First, we develop a cascading failure framework (CFS) using MATPOWER [79], a widely used power-flow simulator and generate synthetic cascading failure data using the IEEE 118-bus topology. Second, a comparison using different classifiers is shown to evaluate the classification performances. The objective is to do exploratory data analysis on labeled data using

various supervised machine learning algorithms and identify the best algorithm based on accuracy. Third, we use a linear regression technique to calculate the number of transmission line failures, and the amount of load shed for any given initial condition.

7.1 Power grid operating parameters and model features

Based on power grid simulations and prior works, we identify the following power 49 grid operating parameters that govern the cascading failure dynamics. In our simulation, we use the IEEE 118-bus system (which is a simple approximation of the American Electric Power system (in the U.S. Midwest) [83]) as the test case which contains 186 transmission lines, 118 buses (nodes) and 54 generators.

Power grid loading level, r : We define the power grid loading level, $r \in [0, 1]$ as the ratio of the total load demand, and the generation capacity of the power grid. In the IEEE 118-bus system, the maximum generation is 9966MW. $r = 1$ indicates the demand is 9966MW and $r \in [0, 1]$ scales the power demand with respect to maximum possible generation. In our simulation, we define a vector $r = \{0.5, 0.6, 0.7, 0.8, 0.9\}$ and simulate the grid against various r . Note that a higher value of r increases the stress in the grid. We observe that for $r < 0.5$, the power grid is under no stress and can absorb the impact of two transmission line failures and redistribute the power flow without any further failures.

Load-shedding constraint, θ : The load-shedding constraint is defined as the ratio of uncontrollable loads (loads that do not participate in load shedding) and the total load in the power grid denoted by, $\theta \in [0, 1]$. This is an important parameter to ensure the control actions by the power grid operator. $\theta = 1$ indicates that all the loads are uncontrollable, and the human operators can perform no load shedding. Again, $\theta = 0$ indicates that the operators can shed any load on the grid. In this chapter, we consider equal load shedding constraints over all the loads in the grid. Further, we consider a vector $\theta = \{0.05, 0.1, 0.15, 0.2, 0.25\}$ and choose

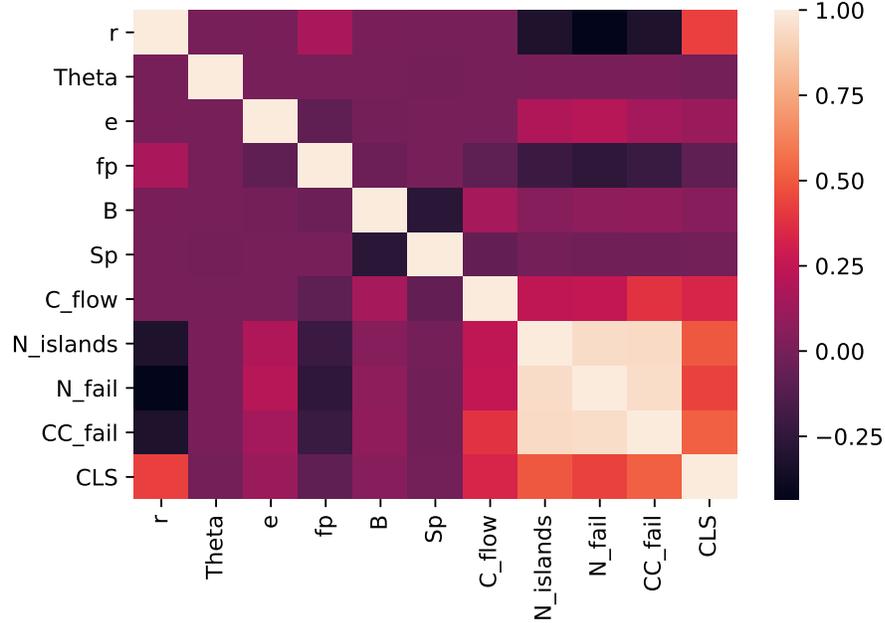


Figure 7.1: Correlation among the features

value of θ randomly from the vector. Similar to r , a higher value of θ increases the probability of cascading failure in the power grid.

Capacity estimation error, e : The Capacity estimation error, $e \in [0, 0.25]$ is defined as the error by the control center in its estimation of the actual capacity of the lines. In our CFS framework, this parameter is used to calculate overloaded lines. We used the same approach used in [5] to calculate overloaded lines. When power flow in a transmission line exceeds $(1-e) \times \text{capacity}$, we consider that line as an overloaded line. We estimate the capacity of a transmission line using power flow simulation under maximum loads, i.e., when generation equals demand ($r=1$). Note that, since we use DC power flow simulation, there are no transient effects, and we can use the maximum generation without any issues. We quantize the flow capacity of a transmission line into a set of five capacities $\{20, 80, 200, 500, 800\}$ MW [81], and assign this capacity of the transmission line as a constraint of the MATPOWER power flow optimization problem (discussed later). In this chapter, we collect cascading failure data using various values of e .

Fixed failure probability of neighboring lines, f_p : To include the effect of hidden failures and localized failures [13], we introduce a parameter namely, the fixed failure probability of neighboring lines, f_p in our CFS framework that fails the first layer adjacent lines of a failed line with a small probability. We consider a vector, $f_p = \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06\}$ and choose the value of f_p randomly from that vector. Since this is a probability, it adds uncertainty on line failures, i.e., the total number of transmission line failure after a cascading failure ends is not deterministic for a given initial condition because this parameter ensures that there is a small equal probability of adjacent line failures for any specific line failures.

Edge betweenness, B and average shortest path, S_p : We keep track of the average edge betweenness of the initially failed lines as a model feature, which is defined as a measure of centrality based on shortest graph distance [91]. Since we are failing two transmission lines initially, we take the average of the edge betweenness as a model feature. Additionally, we track the average shortest path between two of the initially failed transmission lines as a model feature. The shortest path is calculated using Dijkstra's algorithm [91]. In this chapter, to obtain the average shortest path between the two transmission lines, first, we calculate the distance between two starting buses (from the bus) and the two ending bus (to the bus) and then take the average distance between them. The rationale for tracking these two features is to capture the role of power grid physical topology, although, in [7], the authors mentioned that the role of power grid topology in cascading failures in power grids is not explicit due to the dynamics of power flow.

Flow capacity of the initially failed lines, C_{flow} : We keep track of the sum of the flow capacities of the initially failed lines. Intuitively, failing transmission lines with higher capacity yields more transmission line failures in the successive

stages due to the sudden difference between load and generation.

Cumulative installed capacity of the failed lines, CC_{fail} : We keep track of the cumulative installed capacity of the failed lines. Note that, installed capacity of a transmission line is the quantized capacity chosen from the set of five capacities.

Number of Islands, $N_{islands}$: Finally, we include the number of islands formed due to cascading failures in power grids as model parameters. Since islands are formed as a result of failures of transmission lines that breaks the power grid into small self-sufficient microgrids, it is intuitive that the probability of a large cascade is very high if the number of islands is very high.

On top of these features, we track the following two output labels.

Number of failed lines, N_{fail} : We track the number of failed transmission lines after the cascade ends as an output label. We classify the number of failed lines, into three distinct classes (no, small, and large cascade) for classification.

Amount of Load shed, CLS : We use the optimal power flow algorithm from MATPOWER, which includes the capability of implementing load shedding depending on cost. In this work, we set the cost of load shedding ten times higher than the cost of generation to ensure maximum generation before any load shedding. We track the cumulative amount of load shedding as a critical grid parameter.

In general, the stress of the power grid increases as we increase the operating parameters. From the simulations, we observe that depending on the topology, power grid operating parameters, and initial disturbances, the severity of the cascading failure varies from no cascading failure to a complete blackout of the power grid. The correlation among the features is shown in Figure 7.1. Observe that the correlation among the topological parameters B and S_p with the number of failed lines and the amount of load shed is very less. The correlation plot is handy to visualize the correlation between the features.

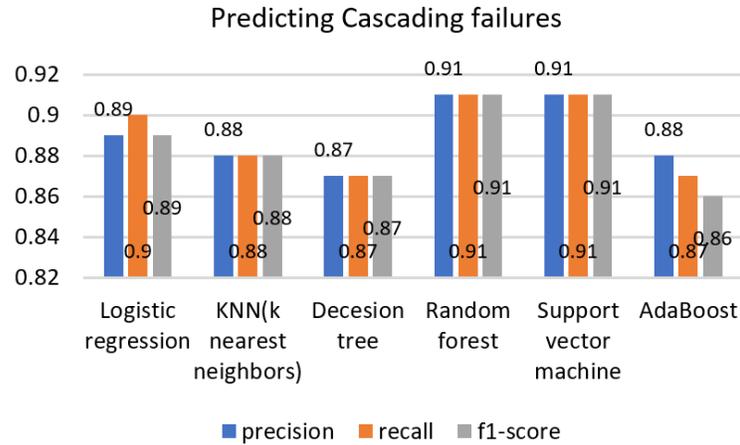


Figure 7.2: Comparison of the overall accuracies for predicting cascading failure in power grids using machine learning algorithms

7.2 Results

In this section, first, we discuss the statistics of the dataset and then implement machine learning algorithms to predict cascading failures in power grids.

7.2.1 Description of the data

Based on our CFS framework, we have performed over seventy-six thousand iterations using the IEEE 118-bus test case to collect synthetic data. In each iteration, we randomly fail two transmission lines. We randomly select r, e, θ, p_n from the vectors defined earlier. We also calculate the topological parameters such as edge betweenness centrality, shortest distance as defined. We store the power flow through the initially failed transmission lines in a column. CFS framework output provides us the number of transmission line failures after cascade ends, the amount of load served, and load shed, number of islands formed, etc. We calculate and store the input-output parameters for all the iterations.

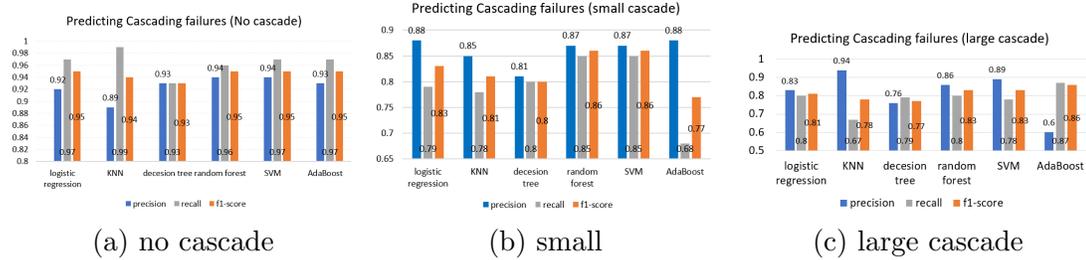


Figure 7.3: Comparison of the accuracies for predicting cascading failure in power grids using machine learning algorithms for the three classes of cascades.

7.2.2 Analysis on cascading failure prediction

We have used the dataset obtained from MATPOWER simulation to perform classification using the python Scikit-learn library [92]. In the dataset, we contain the number of transmission line failures after the cascade ends. As mentioned above, we quantize the number of failed transmission lines in three classes: no cascade (number of transmission lines failures ≤ 10), small cascade ($10 < \text{number of transmission lines failures} < 25$), and large cascade (number of transmission lines failures ≥ 25). We split the dataset for training (70%) and testing (30%) purposes. We have used nine features, as described above. We have used the following machine learning classification algorithms [93, 94]: logistic regression, k-nearest neighbors (KNN), decision tree, random forest, support vector machine (SVM), AdaBoost. The precision, recall, and f1-scores are calculated using [92] for all the algorithms and shown in Figure 7.2. It can be observed that all the classification algorithms have a relatively higher accuracy of classification with SVM and random forest having the best precision. Next, we show the individual cascade type classification accuracies in Figure 7.3. We can observe that the classification of no cascade has higher precision compared to the classification of high cascades. This is because high cascades have low test samples compared to no cascades. For KNN, we further calculate the optimal k that yields the lowest error rate and observed that $k = 9$ gives the highest accuracy.

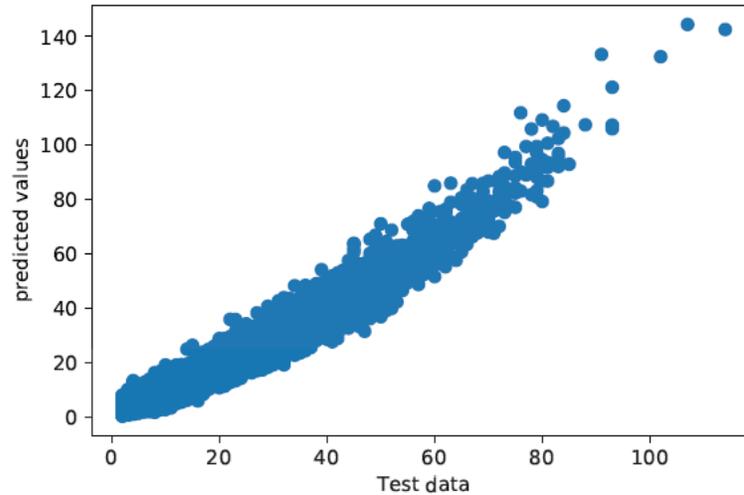


Figure 7.4: Predicting the number of line failures using Linear Regression

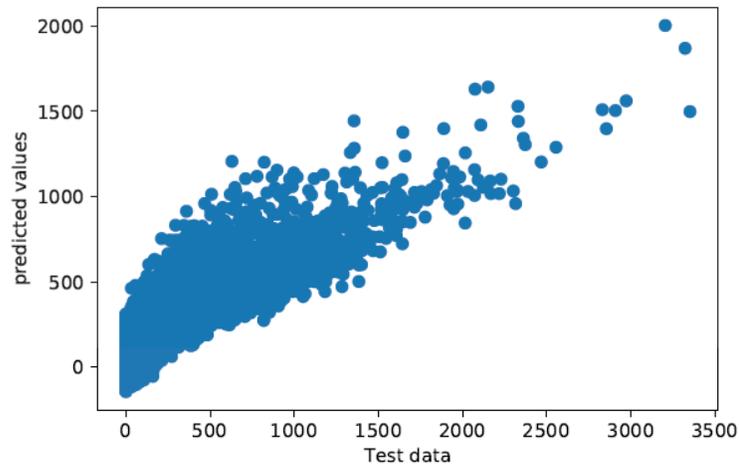


Figure 7.5: Predicting the cumulative amount of load shed using Linear Regression

7.2.3 Linear regression to predict the number of transmission line failures and the amount of load shed

We then use linear regression [93] to predict the number of cascading failure and the amount of load shed which is shown in Figures 7.4 and 7.5 respectively. Observe that, the scatter plot of Figure 7.4 shows that the relationship between the test data and the predicted values are linear, which indicates that linear regression is a reasonable model to predict the number of line failures. The error(deviation)

Table 7.1: Prediction error

Metric	Error (number of failed lines)	Error (amount of load shed)
mean absolute error	1.9	82.92
mean square error	7.07	15334.84
root mean square error	2.66	123.83

of the predicted values from the actual value is reported in Table 7.1. Note that the error is relatively small. However, from Table 7.1, we also observe that the error for predicting the amount of load-shed is relatively large. Also, in Figure 7.5, it is visible that the plot is not linear, which indicates that linear regression is not a good model for predicting the amount of load shedding. Note that, some of the analytical models [5] also predicts the distribution of the number of line failures from simulated data, while in our work, we only predict the number of line failures and the amount of load shed from given data without finding the distribution of the line failures.

7.3 Predicting cascading failures with a new dataset of modified features

We have used the before mentioned cascading failure simulation framework to generate a new dataset with a modified feature. The rationale to generate the new dataset with new features is to prune a few features from the previous dataset and add a few new features to observe the performance. Here it is worth mentioning that most of the features described above are engineered features calculated from the regular features of the smart grid.

7.3.1 Features of the new data set

Following is a brief summary of the features of the new dataset:

Similar to the earlier work, We keep track of the Power grid loading level, r , Load shedding constraint, θ , Capacity estimation error, e .

Failed lines, C_{max} , C_{min} , Installed capacity: We keep track of the initially

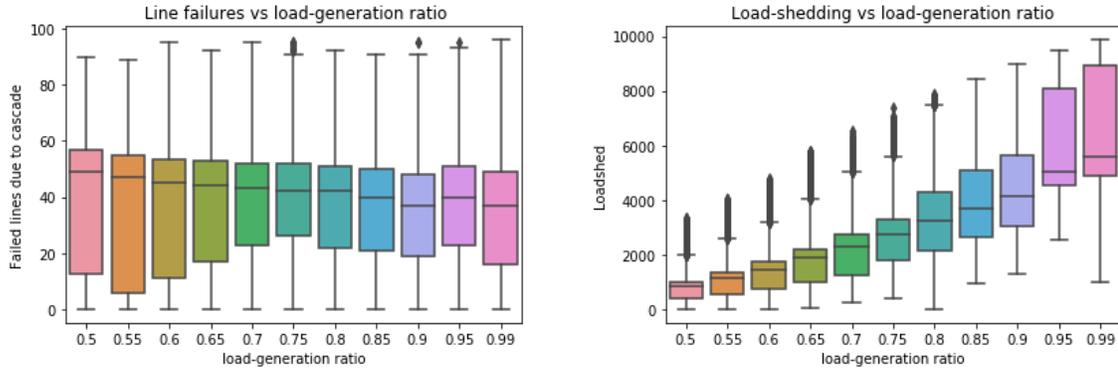


Figure 7.6: box plot of the output variables with different r

failed lines, the maximum, minimum, and cumulative capacity of the initially failed lines as features of the model.

Average degree, average distance: We track the average degree and distance of the network after removing the initially failed lines as topological features of the grid.

Human operator error probability: We use our work in [64] to calculate the human operator error probability, randomly drawn from the distribution of the operator attributes as a feature for the model.

Number of failed lines due to cascade: We track the number of failed transmission lines due to cascade after the cascade ends as an output label.

Amount of load shed: We use the optimal power flow algorithm from MATPOWER, which includes the capability of implementing load shedding depending on the cost. Here, we set the cost of load shedding ten times higher than the cost of generation to ensure maximum generation before any load shedding. We track the cumulative amount of load shedding as a critical grid parameter.

From Figure 7.6, we can see that the number of failed lines almost remains the same as the load increase. However, the amount of loadshed increases significantly, which indicates that the additional load demand is mostly shaded, and fewer line failures are triggered due to this additional load demand. To capture the effect

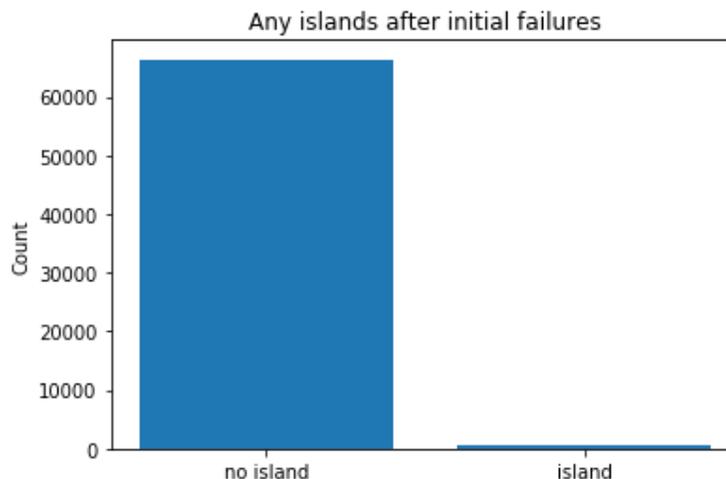


Figure 7.7: Number of islands formed after the initial failures

of cascading failures (triggered by file failures and loadshed), we use the following target variable. Cascading effect: We take the linear combination of the number of failed lines and the amount of load-shed as the output variable we want to predict. The variable is scaled between $[0,1]$. We use the variable values directly for the regression task. For the classification task, we first calculate the median of the cascading effect and use the left half of the median as no cascade (class zero) and the right half as cascade (class one) for classification.

7.3.2 Data cleaning

The dataset was mostly clean, but we checked the following steps to ensure the cleanliness of the data.

- There are no missing values, null values, outliers in the dataset.
- We removed one duplicate column
- We renamed the columns for better understanding

We further checked the dataset with pandas info and describe the method and found everything consistent.

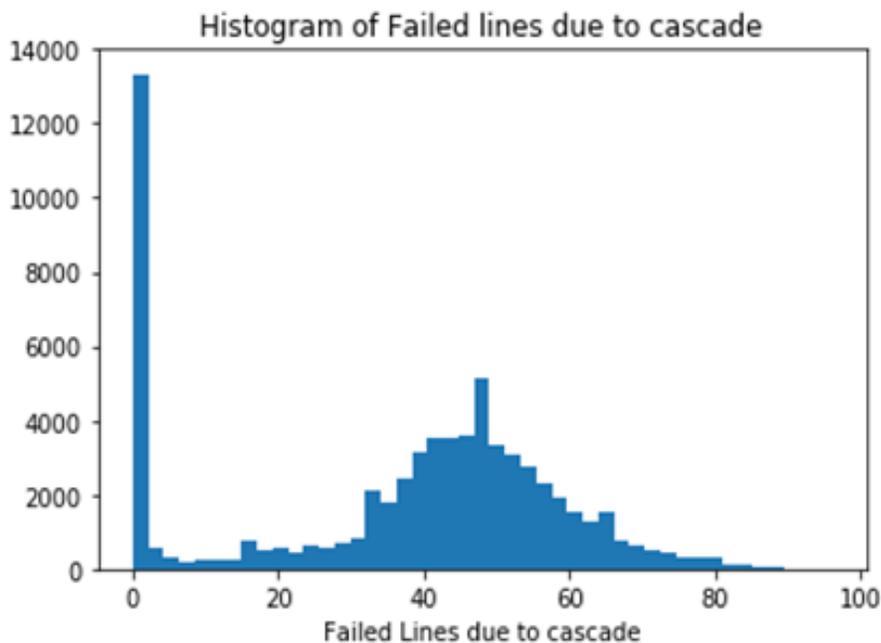


Figure 7.8: Histogram of the failed lines due to cascade

7.3.3 Statistical Analysis

We did statistical analysis such as colinearity analysis for pruning features, correlation analysis, histogram analysis for understanding the behavior of the dataset.

In figure 7.7, we show the number of islands triggered after the initial failed lines of the grid. Since the number of islands triggered is very few, we did not consider it as a feature.

From the histogram of the cascading effect in Figure 5, It can be observed that the histogram is bimodal. The first pick indicates a zone where no additional transmission liens were failed due to cascade, and no loads were shed, i.e., for the set of feature values, no cascading occurred. Similarly, the second pick represents that the average cascading effect occurred at 0.35.

The histogram of load-shedding in Figure 7.9 is slightly skewed to the left, which is intuitive. This indicates the probability of a large cascading failure occurring is less. This also indicates that MATPOWER optimal power flow (OPF)

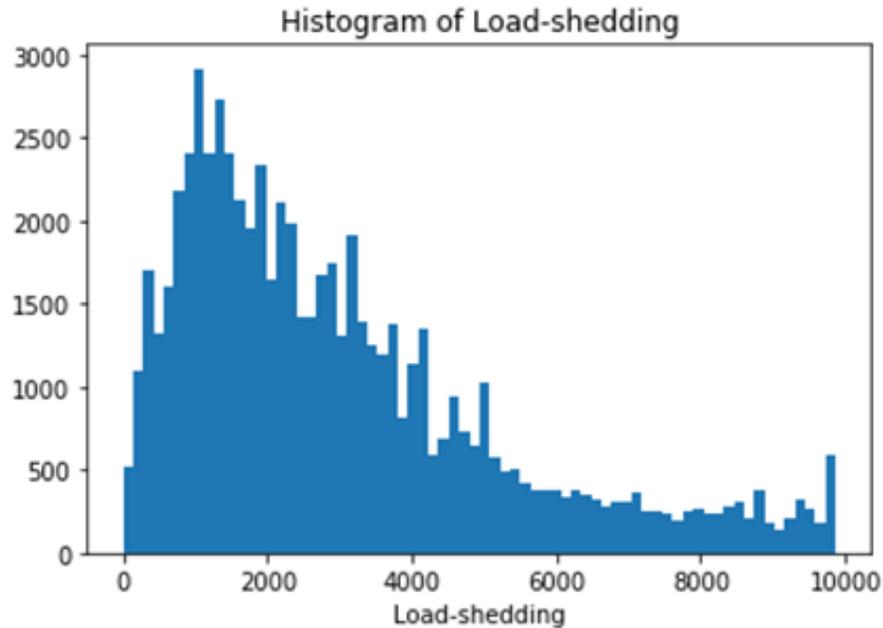


Figure 7.9: Histogram of the load-shedding

is curtailing the loads efficiently to minimize the greater risk of a cascading failure.

From figure 7.10, we can visualize that there is no pattern of correlation between the cascading effect and installed (cumulative) capacity. Also, the installed capacity is not a categorical value. Considering this, we remove this feature from the dataset.

We did not consider generation, served load, and load demand as features of the dataset to avoid collinearity, which can be visualized in Figure 7.11.

We plot the correlation among features in Figure 9. We can observe a strong correlation between the cascading effect and Capacity estimation error. A moderate correlation between cascading effect and load-generation ratio, load-shedding constraint, Human error probability, alpha(negative correlation). Low/minimal Correlation between cascading effect and C_{max} , C_{min} , Degree, distance, and initially failed lines.

From Figure 7.13, the histogram of the cascading effect shows bimodal nature. The first peak is due to no line failure scenarios, and the second peak captures the

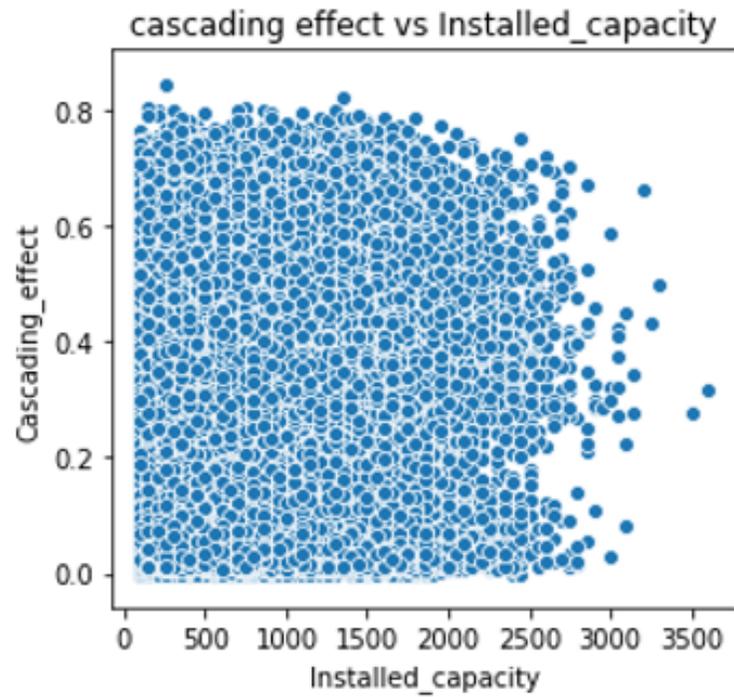


Figure 7.10: Cascading effect vs installed capacity

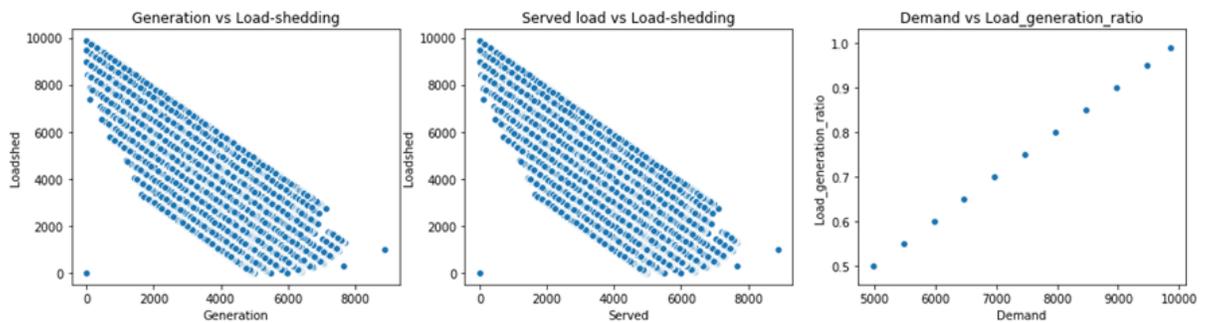


Figure 7.11: Generation and served load vs load shedding, demand vs r

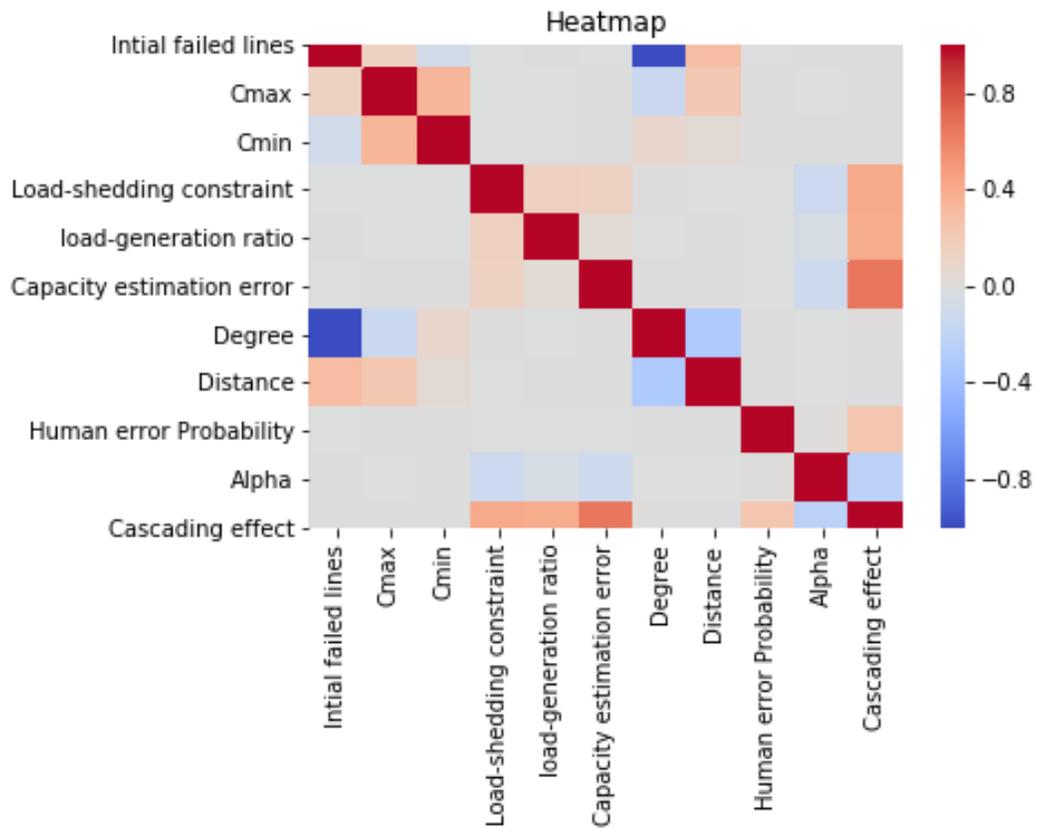


Figure 7.12: Correlation between features for the new dataset

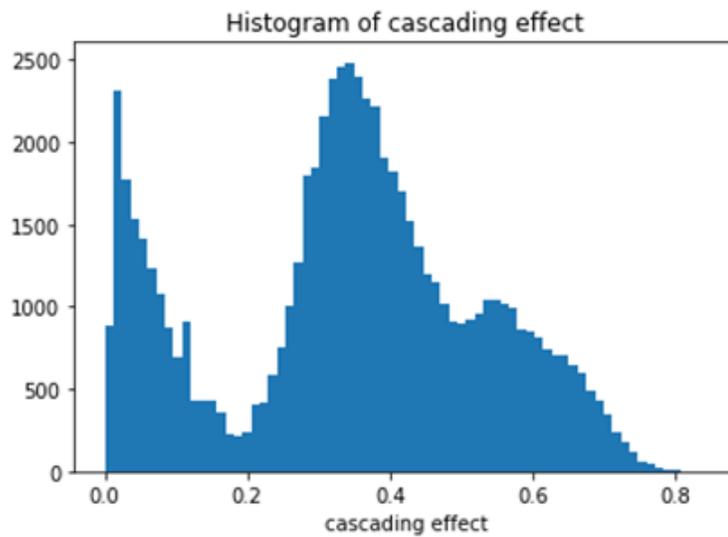


Figure 7.13: Histogram of the cascading effect

average effect of line failures and load shedding.

7.3.4 Results and In-depth analysis using machine learning

We have used the following algorithms for regression:

- Linear regression/ Ridge/Lasso regression
- Random Forest regression
- Support vector regressor

We have used the following algorithms for classification:

- Logistic regression
- KNN (k nearest neighbor)
- Random forest
- Decision tree
- Support vector machine
- Adaboost

The following metrics are used for evaluating the regression performance:

- r-squared score.
- mean absolute error (MAE)
- mean square error (MSE)

The following metrics are used for evaluating the classification performance:

- Accuracy
- Precision

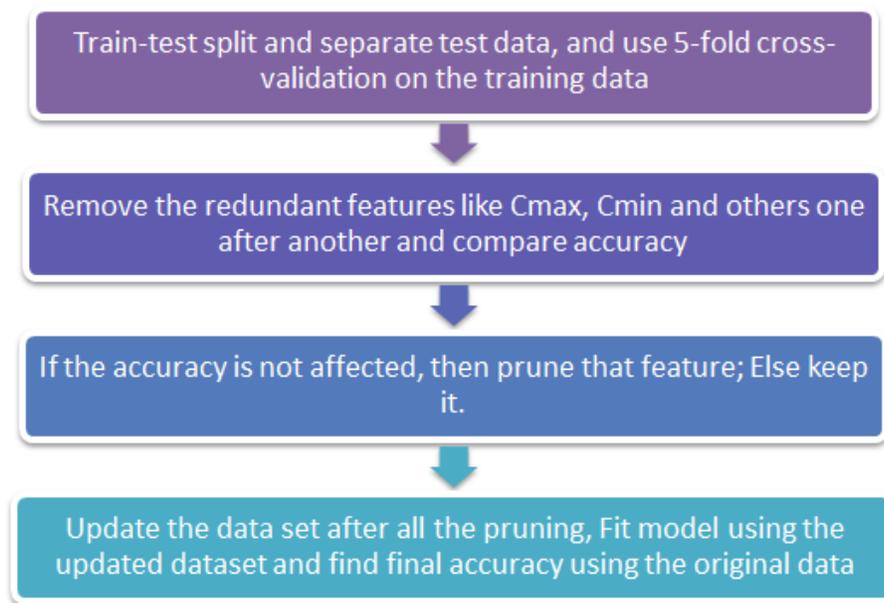


Figure 7.14: Steps used for modeling

- Recall
- F1-score

We show the steps used during the modeling phase in Figure 7.14

We split the data set for training (80%) and testing (20%) purposes. The precision, recall, and f1-scores are calculated using scikit-learn for all the algorithms and shown in Figure 7.15. It can be observed that all the classification algorithms have a relatively higher accuracy of classification with random forest having the best accuracy. Next, we show the individual cascade class classification accuracies in Fig. 12. The purpose of the classification task here is to find the boundary between the cascade and no cascade zones for the given input space. Here also we can see that random forest works best.

The hyperparameters used to get the best accuracy, precision, and recall are given in Table 7.2.

For the regression task, we also obtained the best r-squared error using random forest regression, which is shown in Figure 7.17. The mean squared error is also

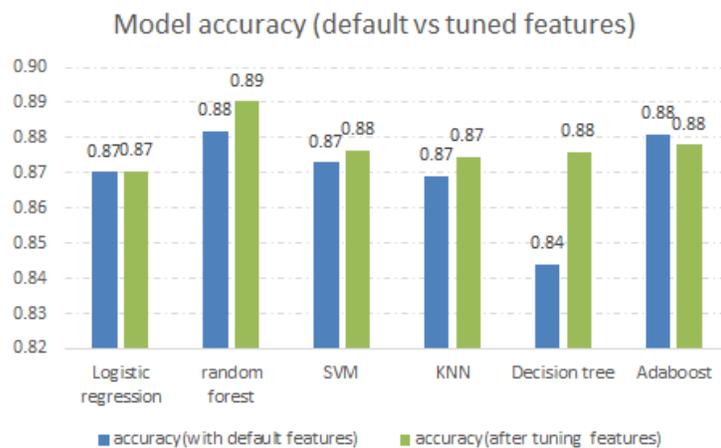


Figure 7.15: Classification accuracy

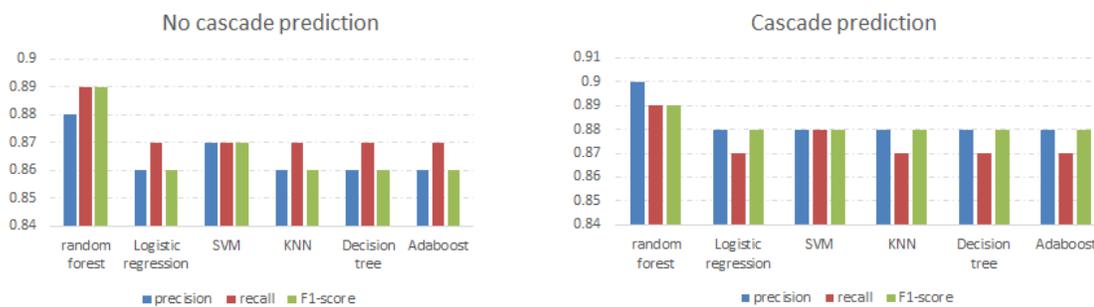


Figure 7.16: Classification accuracy for No cascade/cascade prediction

Table 7.2: Hyperparameters

Model	Hyperparameters
Decision tree	criterion: entropy, min_samples_leaf: 10, min_samples_split: 5
KNN	algorithm: auto, leaf_size: 1, n_neighbors: 10, weights: distance
Adaboost	algorithm: SAMME.R, learning_rate: 0.5, n_estimators: 200
Random forest	criterion: gini, min_samples_leaf: 5, min_samples_split: 5, n_estimators: 50
Logistic regression	C: 10 (penalty = 'l2')
SVM	C: 5, kernel: rbf

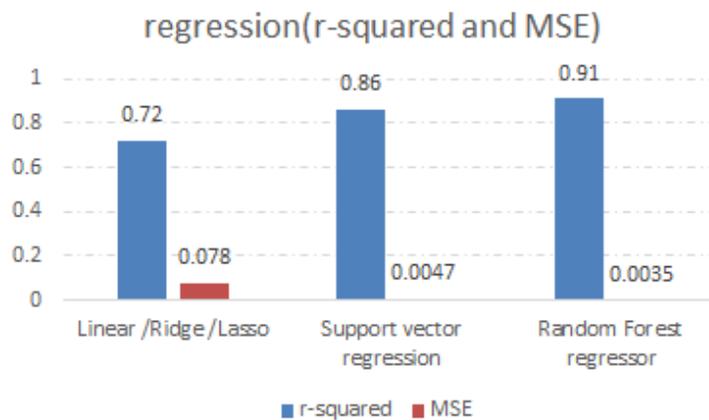


Figure 7.17: Regression on cascading effect

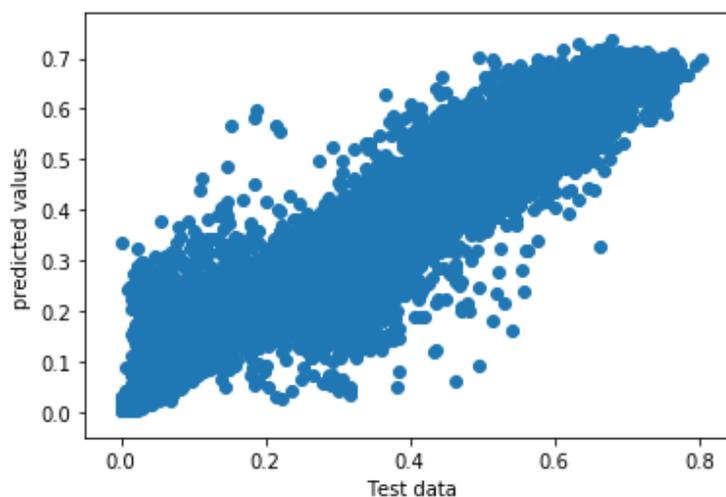


Figure 7.18: Prediction and test data

reported.

Figure 7.18 represents the trend of the linear trend between predicted vs. test values.

Comparing the performance of various models, we select a random forest model for regression and classification tasks. Finally, we have generated 10000 data using the cascading failure simulation framework for testing the model, and we have achieved an 89% accuracy for classification and an r-squared score of 0.90.

Chapter 8

Summary and future works

In this dissertation, a cyber threat and system operator error aware Markov chain based model is presented for analytically predicting cascading failures in the smart grid. Such a model is a significant enhancement and more realistic compared to the existing models for analyzing cascading failures probabilistically. During the propagation of cascading failures, the capability of the system operators, and the reliability of the available information can impact the dynamics of cascading failures significantly, which has been captured effectively in this model. Both benefits and harm of interdependency between power grid layers are captured through the dynamic load-shedding parameter and the cyber threat variable. System operators' error in taking the right actions at the right time is captured through load shedding constraints. Using this I-SASE model, statistics such as the distribution of transmission line failures conditional on initial condition in the steady-state, expected values of the transmission line failures can be predicted, which can be useful to utilities for designing the grid. Additionally, an optimal power-communication inter-connectivity level given initial conditions of the grid can be calculated, which also can be a useful measure of reliability for the grid.

First, in chapter 2, we showed that while studying the cascading-failures in power grids, it is very crucial to analyze the inter-dependency between power, communication, and human interaction, compared to studying the single non-interacting power grid. We showed that the power grid can be negatively influenced when the power grid is under stress, when the human operator is stressed, and when the communication system has failures. A combination of these interactive systems combined with power grid operating parameters can lead to

catastrophic cascading-failures, which are not captured in non-interacting power grid models. We showed that the blackout size in the power grid from a small initial failure could be significantly impacted when the communication network and control operation is not normal. We conclude that under nominal stress level, power grid behaves as a reliable (follows exponential distribution) system during cascading failure even with dependencies from human or communication network; however, if we consider interdependencies from both communication and human operators' and increase stress level, power grid becomes unreliable (follows power-law distribution).

In chapter 3, we have refined a previously-reported Markov chain model to enhance capturing the correlation between the operators' performance attributes with key grid-states, e.g., the number of failed transmission lines and the maximum capacity of the failed lines. The HEPs are determined from eight PSFs. Next, the mapping is created to generate a HEP in terms of the grid states and is embedded into the Markov chain transition matrix to capture the role of operator error using the distribution of the PSFs. Next, a set of critical PSF level combinations are identified using the distribution of the PSF levels that have a high probability of occurrence. Then, the blackout size, including the human error conditional on initial grid conditions, is estimated, and a comparison of results with the existing hSASE model is shown. The use of the distribution of the PSFs is that it allows the capture of the detailed role of the human operator into the cascading failure dynamics. This work is valuable to understand the role of grid operator performance and the impact of operators' error on the reliability of smart grids.

Then, in chapter 4, we present a stochastic Markov-chain based model (I-SASE model) that captures the dynamics of cascading failures in the power grid, including the role of power-communication interdependency and human operator

error. The study of cascading-failure dynamics in a coupled environment, including power, communication, and human operator response, significantly enhances the understanding of cascading-failure dynamics. A combination of these interactive layers, along with grid parameters, can lead to catastrophic cascading-failures which are not captured in single-layered power grid models. The benefits of interdependency are captured by having a higher capability of implementing load shedding during contingencies. On the other hand, strong interdependency increases the probability of cyber-attacks through communication channels and can affect the operators' decision-making capability. Particularly, with the I-SASE model, one can calculate the probability distribution of the number of transmission-line failures in the steady-state as well as can measure the effect of operator error on the propagation of cascading failures. I-SASE model shows that for every fixed grid operating conditions, there exists an optimal level of interdependency between the power grid and the supporting communication infrastructure. The power grid becomes less resilient to contingencies above and under the optimal interdependency as opposed to similar models that suggest that maximum power-communication interdependency ensures maximum resilience to cascading failures.

In chapter 5, a novel way to calculate the average transmission capacity loss (ATCL) analytically during a cascading failure is shown. The model captures the distribution of the transmission-line failures at the steady-state as well as the total transmission-capacity loss. We find a linear correlation between the cumulative capacity of the failed transmission lines with the amount of load-shedding (load loss) is shown. Our model is used to infer the amount of load shed from the ATCL. The model incorporates some operational attributes of the power grid, such as the ratio of the load and the total generation capacity, a constraint on implementing load-shedding and error in estimating the true capacity flow in the transmission lines, which makes the model useful from a practical perspective

by power engineers. Notably, with this model, we can calculate the probability distribution of the number of transmission-line failures and the total capacity of the failed transmission lines as a function of time and at the steady-state. This formulation allows us to identify the operating characteristics as well as the initial conditions (initial line failures and their capacities) that result in cascading failures.

In chapter 6, we analyze the impact of the initial stressor event that leads to cascading failures. We have formulated the initial failures in the power grid with various attack types (Gaussian, circular and linear) and simulate using IEEE 118-bus and 300-bus topology. Our simulations suggest that the number of initially failed transmission lines is linearly proportional to attack intensity. We observe that cascading failures in the power grid are correlated with different power grid parameters during an initial stressor(s) event. These parameters include transmission line failures, the capacity of the failed transmission lines, number of stressor locations, power grid operating parameters such as power grid loading level, load-shedding constrain during an initial stressor event. All these initial conditions eventually determine the blackout-size during a cascading event. Although several models can be found analyzing cascading failures in the power grid, most of them consider arbitrary initial conditions to model the cascading failure behavior. Our work captures the impact of initial conditions during a stressor(s) event and analyzes cascading failures phenomenon from the stressor event that occurred. Future works may include capturing the impact of continuous time-varying degradation functions and identify critical operating settings for the power grid for such degradation functions.

In chapter 7, we have used machine learning algorithms to predict cascading failures in power grids and also used linear regression to predict the number of transmission line failures as well as the cumulative amount of load shedding given

an initial operating condition. First, a CFS framework is developed, which enables us to simulate cascading failure dynamics on the power grid under given initial disturbances (i.e., transmission line failure) and power grid operating setting. Our simulator is built over MATPOWER power flow simulations and runs on any standard IEEE test system case data. Our simulator can effectively generate labeled cascading failure data set under different settings, which is used as an input to machine learning models. Our results suggest that cascading failure prediction can be made using machine learning with high accuracy. However, we showed a simple exploratory data analysis in this work. Complex data-driven modeling to find the distribution of line failures, distribution of load shedding, and critical line identification is missing, which can be a valuable extension of this work. Further, future works include extending the capability of the simulator to include generator dynamics, AC power flow solver, including communication failures, human operator error so that analysis of frequency and transient instability can be analyzed, and smart grid dynamics can be captured.

Future works on continuous improvements of the model and validation in a real grid would be crucial. We have also developed a Markov decision process based on optimal load-shedding policy using the SASE model for mitigating the risk of cascading failures in our research group. Future works would include developing the optimal load-shedding policy using the I-SASE model, capturing the human operators' error and cyber threat. We have used the D.C. optimal power-flow model for calculating the transition probabilities of the Markov chain. Although effective and widely adopted for modeling cascading failures, D.C power flow is a simple approximation of the complex power flow dynamics and omits the effect of transients. Validating the model using A.C. power flow and refining the transition probability formulations to capture the transient effects can be scope for further works in this direction. Further, we have not considered the role of distributed

energy resources (DERs) such as wind farms, solar, and microgrids, as well as operation and management systems. With the significant advent of renewable energy resources in the smart grid, the operation and management of the grid is becoming more and more complex. Inherent intermittent nature of the renewables can significantly change the critical and non-critical operating points of the smart grids. In addition, variable loads during a contingency was not captured in the model. Hence, for completeness of the model, capturing the generator dynamics, variable loads and the role of renewable energy resources into the model would be a significant improvement on enhancing the capability of the model. Further, the model is mainly developed to capture the interdependence between the power grid and the associated communication network in smart grids. The model can be generalized so that analysis of the interdependencies between any interdependent networks (e.g., transportation networks) of any size can be done. One drawback of the model is that due to the unavailability of full grid network simulation data, the analytical model for cyber threat transitions could not be calculated analytically. Also, the cyber security model approximated here is very naive. Detailed modeling efforts are needed to understand the dynamics of cyber attacks on cascading failures more precisely. Future efforts in validating the model in a live testbed would be crucial before implementing the model in a live grid network. Again, works can be done to Understand the role of cyber-threat awareness in the decision process of system operators. With the advent of complex communication capability, AI-based decision support tools are gaining more popularity primarily due to the level of complex calculations and rule-based decisions it can manage in no time to reduce human error. On the other hand, too much dependence on AI-based tools would lead to a catastrophe in unforeseen events. A study is needs to be done to find the optimal level of interaction between system operators and the AI-based automated decision support systems.

BIBLIOGRAPHY

- [1] Zhuoyao Wang, Mahshid Rahnamay-Naeini, Joana M Abreu, Rezoan A Shuvro, Pankaz Das, Andrea A Mammoli, Nasir Ghani, and Majeed M Hayat. Impacts of operators' behavior on reliability of power grids during cascading failures. *IEEE Transactions on Power Systems*, 33(6):6013–6024, 2018.
- [2] List of major power outages. https://en.wikipedia.org/wiki/List_of_major_power_outages. Accessed: 2019-07-15.
- [3] A Muir and J Lopatto. Final report on the August 14, 2003 blackout in the united states and canada: causes and recommendations, 2004.
- [4] JH Eto. Blackout 2003: Final report on the August , 2003 blackout in the united states and canada: causes and recommendations. *Electricity Markets and Policy Group, Energy Analysis and Environmental Impacts Department, US Department of Energy, Washington, DC*, 2004.
- [5] Mahshid Rahnamay-Naeini, Zhuoyao Wang, Nasir Ghani, Andrea Mammoli, and Majeed M Hayat. Stochastic analysis of cascading-failure dynamics in power grids. *IEEE Transactions on Power Systems*, 29(4):1767–1779, 2014.
- [6] Mahshid Rahnamay-Naeini and Majeed M Hayat. On the role of power-grid and communication-system interdependencies on cascading failures. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 527–530. IEEE, 2013.
- [7] Mert Korkali, Jason G Veneman, Brian F Tivnan, James P Bagrow, and Paul DH Hines. Reducing cascading failure risk by increasing infrastructure network interdependence. *Scientific reports*, 7:44499, 2017.
- [8] Pankaz Das, Mahshid Rahnamay-Naeini, Nasir Ghani, and Majeed M Hayat. On the vulnerability of multi-level communication network under catastrophic events. In *2017 International Conference on Computing, Networking and Communications (ICNC)*, pages 912–916. IEEE, 2017.
- [9] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart grid—the new and improved power grid: A survey. *IEEE Communications Surveys and Tutorials*, 14(4):944–980, 2011.
- [10] Kai Sun, Yunhe Hou, Wei Sun, and Junjian Qi. *Power System Control Under Cascading Failures: Understanding, Mitigation, and System Restoration*. Wiley-IEEE Press, 2019.
- [11] Ross Baldick, Badrul Chowdhury, Ian Dobson, Zhaoyang Dong, Bei Gou, David Hawkins, Henry Huang, Manho Joung, Daniel Kirschen, and Fangxing

- Li. Initial review of methods for cascading failure analysis in electric power transmission systems ieeepes task force on understanding, prediction, mitigation and restoration of cascading failures. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–8. IEEE, 2008.
- [12] Paul Hines, Jay Apt, and Sarosh Talukdar. Trends in the history of large blackouts in the united states. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–8. IEEE, 2008.
- [13] Hengdao Guo, Ciyan Zheng, Herbert Ho-Ching Iu, and Tyrone Fernando. A critical review of cascading failure analysis and modeling of power system. *Renewable and Sustainable Energy Reviews*, 80:9–22, 2017.
- [14] Massoud Amin and Phillip F Schewe. Preventing blackouts. *Scientific American*, 296(5):60–67, 2007.
- [15] Alexander Veremyev, Alexey Sorokin, Vladimir Boginski, and Eduardo L Pasiliao. Minimum vertex cover problem for coupled interdependent networks with cascading failures. *European Journal of Operational Research*, 232(3):499–511, 2014.
- [16] NERC Ferc. Arizona-southern california outages on 8 september 2011: causes and recommendations. *FERC and NERC*, 2012.
- [17] Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016.
- [18] Paolo Crucitti, Vito Latora, and Massimo Marchiori. Model for cascading failures in complex networks. *Physical Review E*, 69(4):045104, 2004.
- [19] Adilson E Motter. Cascade control and defense in complex networks. *Physical Review Letters*, 93(9):098701, 2004.
- [20] Charles D Brummitt, Raissa M D’Souza, and Elizabeth A Leicht. Suppressing cascades of load in interdependent networks. *National Academy of Sciences*, 109(12):E680–E689, 2012.
- [21] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.
- [22] Giuliano Andrea Pagani and Marco Aiello. The power grid as a complex network: a survey. *Physica A: Statistical Mechanics and its Applications*, 392(11):2688–2700, 2013.

- [23] Run-Ran Liu, Ming Li, Chun-Xiao Jia, and Bing-Hong Wang. Cascading failures in coupled networks with both inner-dependency and inter-dependency links. *Scientific reports*, 6:25294, 2016.
- [24] Paolo Crucitti, Vito Latora, and Massimo Marchiori. A topological analysis of the italian electric power grid. *Physica A: Statistical mechanics and its applications*, 338(1-2):92–97, 2004.
- [25] Ryan Kinney, Paolo Crucitti, Reka Albert, and Vito Latora. Modeling cascading failures in the north american power grid. *The European Physical Journal B-Condensed Matter and Complex Systems*, 46(1):101–107, 2005.
- [26] Réka Albert, István Albert, and Gary L Nakarado. Structural vulnerability of the north american power grid. *Physical review E*, 69(2):025103, 2004.
- [27] Jia Shao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Cascade of failures in coupled network systems with multiple support-dependence relations. *Physical Review E*, 83(3):036116, 2011.
- [28] EA Leicht and Raissa M D’Souza. Percolation on interacting networks. *arXiv preprint arXiv:0907.0894*, 2009.
- [29] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Non-linear Science*, 17(2):026103, 2007.
- [30] Marzieh Parandehgheibi, Eytan Modiano, and David Hay. Mitigating cascading failures in interdependent power grids and communication networks. In *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 242–247. IEEE, 2014.
- [31] Benjamin A Carreras, David E Newman, Paul Gradney, Vickie E Lynch, and Ian Dobson. Interdependent risk in interacting infrastructure systems. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, pages 112–112. IEEE, 2007.
- [32] Yang Yang, Takashi Nishikawa, and Adilson E Motter. Small vulnerable sets determine large network cascades in power grids. *Science*, 358(6365):eaan3184, 2017.
- [33] Benjamin Schäfer, Dirk Witthaut, Marc Timme, and Vito Latora. Dynamically induced cascading failures in power grids. *Nature communications*, 9(1):1–13, 2018.
- [34] Andrey Bernstein, Daniel Bienstock, David Hay, Meric Uzunoglu, and Gil Zussman. Power grid vulnerability to geographically correlated failures—analysis and control implications. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 2634–2642. IEEE, 2014.

- [35] Dimitris Achlioptas, Raissa M D'Souza, and Joel Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009.
- [36] Osman Yagan, Dajun Qian, Junshan Zhang, and Douglas Cochran. Optimal allocation of interconnecting links in cyber-physical systems: Interdependence, cascading failures, and robustness. *IEEE Transactions on Parallel and Distributed Systems*, 23(9):1708–1720, 2012.
- [37] Saleh Soltan, Dorian Mazaauric, and Gil Zussman. Analysis of failures in power grids. *IEEE Transactions on Control of Network Systems*, 4(2):288–300, 2015.
- [38] Saleh Soltan, Alexander Loh, and Gil Zussman. A learning-based method for generating synthetic power grids. *IEEE Systems Journal*, 13(1):625–634, 2018.
- [39] Hale Cetinay, Saleh Soltan, Fernando A Kuipers, Gil Zussman, and Piet Van Mieghem. Analyzing cascading failures in power grids under the ac and dc power flow models. *ACM SIGMETRICS Performance Evaluation Review*, 45(3):198–203, 2018.
- [40] Saleh Soltan and Gil Zussman. Generation of synthetic spatially embedded power grid networks. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2016.
- [41] Benjamin A Carreras, Vickie E Lynch, Ian Dobson, and David E Newman. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos: An interdisciplinary journal of nonlinear science*, 12(4):985–994, 2002.
- [42] Hua Lin, Santhoshkumar Sambamoorthy, Sandeep Shukla, James Thorp, and Lamine Mili. Power system and communication network co-simulation for smart grid applications. In *ISGT 2011*, pages 1–6. IEEE, 2011.
- [43] Kenneth Hopkinson, Xiaoru Wang, Renan Giovanini, James Thorp, Kenneth Birman, and Denis Coury. Epochs: a platform for agent-based electric power and communication simulation built from commercial off-the-shelf components. *IEEE Transactions on Power Systems*, 21(2):548–558, 2006.
- [44] Mahshid Rahnamay-Naeini and Majeed M Hayat. Cascading failures in interdependent infrastructures: An interdependent markov-chain approach. *IEEE Transactions on Smart Grid*, 7(4):1997–2006, 2016.
- [45] Zhifang Wang, Anna Scaglione, and Robert J Thomas. A markov-transition model for cascading failures in power grids. In *2012 45th Hawaii International Conference on System Sciences*, pages 2115–2124. IEEE, 2012.
- [46] Marian Anghel, Kenneth A Werley, and Adilson E Motter. Stochastic model for power grid dynamics. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pages 113–113. IEEE, 2007.

- [47] Christopher L DeMarco. A phase transition model for cascading network failure. *IEEE control systems*, 21(6):40–51, 2001.
- [48] Paul Hines, Eduardo Cotilla-Sanchez, and Seth Blumsack. Topological models and critical slowing down: Two approaches to power system blackout risk analysis. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2011.
- [49] Paul Hines, Karthikeyan Balasubramaniam, and Eduardo Cotilla Sanchez. Cascading failures in power grids. *Ieee Potentials*, 28(5):24–30, 2009.
- [50] Hui Ren and Ian 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*, 55(9):927–931, 2008.
- [51] Ian Dobson. Estimating the propagation and extent of cascading line outages from utility data with a branching process. *IEEE Transactions on Power Systems*, 27(4):2146–2155, 2012.
- [52] Ian Dobson, Benjamin A Carreras, and David E Newman. A loading-dependent model of probabilistic cascading failure. *Probability in the Engineering and Informational Sciences*, 19(1):15–32, 2005.
- [53] Sakshi Pahwa, Caterina Scoglio, and Antonio Scala. Abruptness of cascade failures in power grids. *Scientific reports*, 4:3694, 2014.
- [54] B Liscouski and W Elliot. Final report on the August 14, 2003 blackout in the united states and canada: Causes and recommendations. *A report to US Department of Energy*, 40(4):86, 2004.
- [55] Mahshid Rahnamay-Naeini, Zhuoyao Wang, Andrea Mammoli, and Majeed M Hayat. A probabilistic model for the dynamics of cascading failures and blackouts in power grids. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–8. IEEE, 2012.
- [56] Marianna Vaiman, Keith Bell, Yousu Chen, Badrul Chowdhury, Ian Dobson, Paul Hines, Milorad Papic, Stephen Miller, and Pei Zhang. Risk assessment of cascading outages: Methodologies and challenges. *IEEE Transactions on Power Systems*, 27(2):631, 2012.
- [57] Janusz Bialek, Emanuele Ciapessoni, Diego Cirio, Eduardo Cotilla-Sanchez, Chris Dent, Ian Dobson, Pierre Henneaux, Paul Hines, Jorge Jardim, and Stephen Miller. Benchmarking and validation of cascading failure analysis tools. *IEEE Transactions on Power Systems*, 31(6):4887–4900, 2016.
- [58] Pierre Henneaux, Emanuele Ciapessoni, Diego Cirio, Eduardo Cotilla-Sanchez, Ruisheng Diao, Ian Dobson, Anish Gaikwad, Stephen Miller, Milorad Papic, and Andrea Pitto. Benchmarking quasi-steady state cascading

- outage analysis methodologies. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6. IEEE, 2018.
- [59] Pankaz Das, Rezoan A Shuvro, Zhuoyao Wang, Majeed M Hayat, and Francesco Sorrentino. A data-driven model for simulating the evolution of transmission line failure in power grids. In *2018 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2018.
- [60] Priyanka Dey, Rachit Mehra, Faruk Kazi, Sushama Wagh, and Navdeep M Singh. Impact of topology on the propagation of cascading failure in power grid. *IEEE Transactions on Smart Grid*, 7(4):1970–1978, 2016.
- [61] Ian Dobson et al. Branching process models for the exponentially increasing portions of cascading failure blackouts. In *System Sciences, 2005. HICSS’05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 64a–64a. IEEE, 2005.
- [62] Rezoan A Shuvro, Zhuoyao Wang, Pankaz Das, Mahshid R Naeini, and Majeed M Hayat. Modeling impact of communication network failures on power grid reliability. In *North American Power Symposium*. IEEE, 2017.
- [63] David Gertman, Harold Blackman, Julie Marble, James Byers, and Curtis Smith. The spar-h human reliability analysis method. *US Nuclear Regulatory Commission*, 230:35, 2005.
- [64] Joana M Abreu, Natasha Hardy, Francisco C Pereira, and Michael Zeifman. Modeling human reliability in the power grid environment: An application of the spar-h methodology. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 662–666. SAGE Publications Sage CA: Los Angeles, CA, 2015.
- [65] Ian Dobson. Where is the edge for cascading failure?: challenges and opportunities for quantifying blackout risk. In *Power Engineering Society General Meeting, 2007. IEEE*, pages 1–8. IEEE, 2007.
- [66] Ronald L Boring and Harold S Blackman. The origins of the spar-h methods performance shaping factor multipliers. In *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting*, pages 177–184. IEEE, 2007.
- [67] Harold S Blackman, David I Gertman, and Ronald L Boring. Human error quantification using performance shaping factors in the spar-h method. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 52, pages 1733–1737. SAGE Publications Sage CA: Los Angeles, CA, 2008.

- [68] Rezoan A Shuvro, Zhuoyao Wangt, Pankaz Das, Mahshid R Naeini, and Majeed M Hayat. Modeling cascading failures in power grids including communication and human operator impacts. In *2017 IEEE Green Energy and Smart Systems Conference (IGESSC)*.
- [69] Sudha Gupta, Ruta Kambli, Sushama Wagh, and Faruk Kazi. Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. *IEEE Transactions on Industrial Electronics*, 62(4):2478–2486, 2014.
- [70] Renjian Pi, Ye Cai, Yong Li, and Yijia Cao. Machine learning based on bayes networks to predict the cascading failure propagation. *IEEE Access*, 6:44815–44823, 2018.
- [71] Raymond C Borges Hink, Justin M Beaver, Mark A Buckner, Tommy Morris, Uttam Adhikari, and Shengyi Pan. Machine learning for power system disturbance and cyber-attack discrimination. In *2014 7th international symposium on resilient control systems (ISRCS)*, pages 1–8. IEEE, 2014.
- [72] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [73] Abdulaziz Almalaq and George Edwards. A review of deep learning methods applied on load forecasting. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 511–516. IEEE, 2017.
- [74] Wenye Wang, Yi Xu, and Mohit Khanna. A survey on the communication architectures in smart grid. *Computer networks*, 55(15):3604–3629, 2011.
- [75] Pankaz Das, Rezoan A Shuvro, Zhuoyao Wang, Mahshid Rahnamay Naeini, Nasir Ghani, and Majeed M Hayat. Stochastic failure dynamics in communication network under the influence of power failure. In *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–8. IEEE, 2017.
- [76] Rezoan A Shuvro, Zhuoyao Wang, Pankaz Das, Mahshid R Naeini, and Majeed M Hayat. Modeling impact of communication network failures on power grid reliability. In *2017 North American Power Symposium (NAPS)*, pages 1–6. IEEE, 2017.
- [77] Mahshid Rahnamay-Naeini, Zhuoyao Wang, Andrea Mammoli, and Majeed M Hayat. Impacts of control and communication system vulnerabilities on power systems under contingencies. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–7. IEEE, 2012.
- [78] Yeong-Taeg Kim. Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE transactions on Consumer Electronics*, 43(1):1–8, 1997.

- [79] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26(1):12–19, 2010.
- [80] Mahshid Rahnamay Naeini. *Stochastic dynamics of cascading failures in electric-cyber infrastructures*. The University of New Mexico, 2014.
- [81] Allen J Wood, Bruce F Wollenberg, and Gerald B Sheblé. *Power generation, operation, and control*. John Wiley & Sons, 2013.
- [82] Christopher Z Mooney. *Monte carlo simulation*, volume 116. Sage publications, 1997.
- [83] Rich Christie. Ieee 118 bus test case. *College of Engineering, Electric Engineering, University of Washington*, 1993.
- [84] Quan Chen and Lamine Mili. Composite power system vulnerability evaluation to cascading failures using importance sampling and antithetic variates. *IEEE transactions on power systems*, 28(3):2321–2330, 2013.
- [85] Zhuoyao Wang, Mahshid Rahnamay-Naeini, Joana M Abreu, Rezoan A Shuvro, Das Pankaz, Andrea Mammoli, Nasir Ghani, and Majeed M Hayat. Modeling and analyzing impacts of operators behavior on infrastructure reliability during contingencies. In *article can be found on this url <http://ece-research.unm.edu/sg/wp-content/uploads/2017/06/hSASE2017.pdf>*.
- [86] Spar-h step-by-step guidance. <https://www.nrc.gov/docs/ML1120/ML112060305.pdf>. Accessed: 2020-04-22.
- [87] Rezoan A Shuvro, Pankaz Das, Joana M Abreu, and Majeed M Hayat. Correlating grid-operators’ performance with cascading failures in smart-grids. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pages 1–5. IEEE.
- [88] Rezoan A Shuvro, Pankaz Das, Joana M Abreu, and Majeed M Hayat. Correlating grid-operators’ performance with cascading failures in smart-grids. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pages 1–5. IEEE.
- [89] Pankaz Das, Rezoan A Shuvro, Zhuoyao Wang, Majeed M Hayat, and Francesco Sorrentino. A data-driven model for simulating the evolution of transmission line failure in power grids. In *IEEE North American Power Symposium*, 2018.
- [90] Mahshid Rahnamay-Naeini, Jorge E Pezoa, Ghady Azar, Nasir Ghani, and Majeed M Hayat. Modeling stochastic correlated failures and their effects on network reliability. In *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–6. IEEE, 2011.

- [91] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.
- [92] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [93] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [94] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24, 2007.