Summer 2017

Quantitative dependability and interdependency models for large-scale cyber-physical systems

Koosha Marashi

QUANTITATIVE DEPENDABILITY AND INTERDEPENDENCY MODELS FOR LARGE-SCALE CYBER-PHYSICAL SYSTEMS

by

KOOSHA MARASHI

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

2017

Approved by

Sahra Sedigh Sarvestani, Co-Advisor
Ali R. Hurson, Co-Advisor
Minsu Choi
Joe Stanley
Maciej J. Zawodniok

ABSTRACT

Cyber-physical systems link cyber infrastructure with physical processes through an integrated network of physical components, sensors, actuators, and computers that are interconnected by communication links. Modern critical infrastructures such as smart grids, intelligent water distribution networks, and intelligent transportation systems are prominent examples of cyber-physical systems. Developed countries are entirely reliant on these critical infrastructures, hence the need for rigorous assessment of the trustworthiness of these systems. The objective of this research is quantitative modeling of dependability attributes - including reliability and survivability - of cyber-physical systems, with domain-specific case studies on smart grids and intelligent water distribution networks. To this end, we make the following research contributions: i) quantifying, in terms of loss of reliability and survivability, the effect of introducing computing and communication technologies; and ii) identifying and quantifying interdependencies in cyber-physical systems and investigating their effect on fault propagation paths and degradation of dependability attributes.

Our proposed approach relies on observation of system behavior in response to disruptive events. We utilize a Markovian technique to formalize a unified reliability model. For survivability evaluation, we capture temporal changes to a service index chosen to represent the extent of functionality retained. In modeling of interdependency, we apply correlation and causation analyses to identify links and use graph-theoretical metrics for quantifying them. The metrics and models we propose can be instrumental in guiding investments in fortification of and failure mitigation for critical infrastructures.To verify the success of our proposed approach in meeting these goals, we introduce a failure prediction tool capable of identifying system components that are prone to failure as a result of a specific disruptive event. Our prediction tool can enable timely preventative actions and mitigate the consequences of accidental failures and malicious attacks.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Dr. Sahra Sedigh Sarvestani and Dr. Ali Hurson, for their guidance and support over the past years. Dr. Sedigh has all the qualities a great advisor should have. She will be my role model for many aspects of my life. I also thank Dr. Hurson for giving me the insight to approach intricate problems and sharing his valuable experiences with me. I would like to extend my appreciation to Dr. Minsu Choi, Dr. Joe Stanley, and Dr. Maciej Zawodniok for serving on my dissertation advisory committee and for reviewing this dissertation and giving constructive feedback.

I want to express my gratitude to my collaborators in the SeNDeComp research lab. Working with them was a great learning experience. I must also thank my dear friends, with whom I shared so many unforgettable memories, Mazdak Memarnezhad, Amin Sadeghizadeh, Navid Rezazadeh, Ardalan Hosseini, Behzad Amiri, Sajjad Amini, Amir Ghazanfari, Reza Rahimi, Peyman Heidari, Alireza Toghraee, Mojtaba Ale Mohammadi, Hesam Zomorodi, and Maryam Kazemzadeh.

I would like to thank my parents for their constant love, support, and faith in me throughout my life. My parents raised and selflessly provided me with all I needed to become a successful individual. I also thank my lovely brother and sister, Pouya and Paria, whose unwavering support have kept me going during the ups and downs of my life.

Last, but certainly not least, is my girlfriend, Baharan. I cannot find words to express my gratitude to her. She is always supportive of my choices, and often makes sacrifices to enable them. Her constant support and unconditional love made my graduate school career substantially easier and more joyful.

This work was supported in part by the US Departments of Transportation and Education, the National Science Foundation, the Missouri S&T Intelligent Systems Center, and the Center for Infrastructure Engineering Studies.

TABLE OF CONTENTS

Page

SECTION

APPENDICES

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

n	Number of components
m	Number of simulated failure cases
$\mathcal{F}_k(t)$	Set of components that degrade at time t during failure case k
$\mathbf{X}(t)$	Vector of component state variables
PCC	Pearson correlation coefficient
RDC	Randomized dependency coefficient
$\mathbf{W}$	Succession frequency matrix
$\mathcal{H}_{k,j}$	Set of components that cause degradation of component j, in failure case k
$\mathbf{E}$	Causation frequency matrix
f_i	Number of times, among all failure cases, that component i experiences degradation
$\mathbf{D}$	Direct influence matrix
$\mathbf{T}$	Total influence matrix
τ_i	Weighted out-degree of component i
ν_i	Weighted in-degree of component i
$\gamma_{s_1-s_2}$	Interdependency index between subsystems s_1 and s_2
$\mathbf{Y}$	Vector of predicted failure probabilities
C_i	Set of states for component i

N_i	Cardinality of $\mathcal{C}_i$
$\Phi(t)$	System state at time t
$\mathcal{S}$	System state space
N	Cardinality of $\mathcal{S}$
$\mathbf{U}$	Binary vector identifying each system state as failed or operational
$\Pi(t)$	State distribution vector at time t
Λ	State transition probability matrix
Λ_i	State transition probability matrix, based on operation of component i
p_i	Reliability of component i
q_i	Unreliability of component i
R	System reliability
$M_k(t)$	Record of FoM for failure case k
δ_k	Full extent of degradation in failure case k
Δ	Maximum δ over all failure cases
ρ_k	The most rapid rate of degradation in failure case k
$\mathcal{G}_k$	Set of all components failed in failure case k
$\mathcal{Q}_i$	Set of failure cases in which component i has failed
$t_i^{(k)}$	Time at which component i fails during failure case k
α_i	Criticality of component i
β_i	Fragility of component i

1. INTRODUCTION

Modern critical infrastructures are large complex systems that are expected to be highly dependable and continuously provide essential services. Examples of such complex networks are smart grids, intelligent water distribution networks, and intelligent transportation systems. These systems utilize cyber infrastructure, which provides computing-based decision support and failure monitoring, among other benefits. Collectively, the physical infrastructure and this network of cyber devices comprise a cyber-physical system (CPS). A fundamental motivation for supplementing the existing physical systems with computing and communication is to improve dependability. The research presented in this dissertation is devoted to determining whether CPSs succeed in providing a higher level of dependability than their conventional counterparts.

Dependability analysis of CPSs has become increasingly urgent, given the ubiquity of such systems, extensive use of computing in critical applications, and disruptions that inevitably occur in critical infrastructures. Past incidents have proven that incapacity of critical infrastructures can have a catastrophic impact on our health, safety, economics, and social welfare [2, 3, 4, 5, 6].

In the energy sector for example, there has been instances where unexpected failures in the electric delivery systems has resulted in large-scale power outages with profound consequences on several sectors of the critical infrastructure. Figure 1.1 depicts three large-scale catastrophic power outages since 2000 and highlights the main recommendations made by task forces that analyzed respective events.

In 2003 Northeastern Blackout, the outage of transmission lines, combined with a failure in the alarm system, caused an instability and resulted in a series of cascading failures. These incidents eventually led to a large-scale blackout that affected more than 55 million people [2]. The 2003 Italy Blackout was a cascading power failure that left

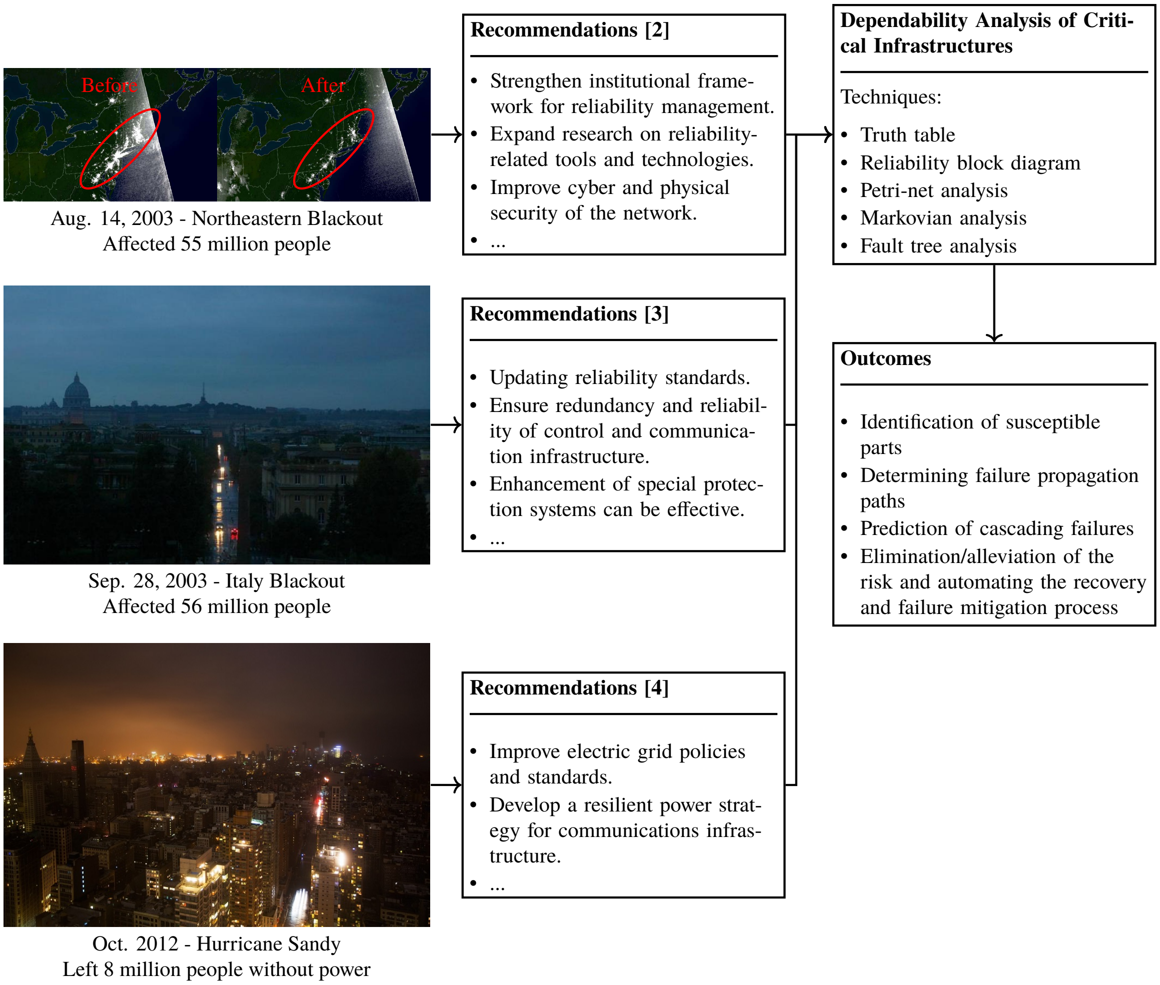

Before
After
Aug. 14, 2003 - Northeastern Blackout
Affected 55 million people
Sep. 28, 2003 - Italy Blackout
Affected 56 million people
Oct. 2012 - Hurricane Sandy
Left 8 million people without power
Recommendations [2]
• Strengthen institutional framework for reliability management.
• Expand research on reliability-related tools and technologies.
• Improve cyber and physical security of the network.
• ...
Recommendations [3]
• Updating reliability standards.
• Ensure redundancy and reliability of control and communication infrastructure.
• Enhancement of special protection systems can be effective.
• ...
Recommendations [4]
• Improve electric grid policies and standards.
• Develop a resilient power strategy for communications infrastructure.
• ...
Dependability Analysis of Critical Infrastructures
Techniques:
• Truth table
• Reliability block diagram
• Petri-net analysis
• Markovian analysis
• Fault tree analysis
Outcomes
• Identification of susceptible parts
• Determining failure propagation paths
• Prediction of cascading failures
• Elimination/alleviation of the risk and automating the recovery and failure mitigation process

Figure 1.1. Catastrophic events that motivate this research.

half of the country without power for multiple days [3]. This failure was exacerbated by the loss of Internet communication nodes due to the power outage, which in turn caused further breakdown of communication and control at multiple power stations. Hurricane Sandy is an example where several critical infrastructures were affected due to a natural disaster [4]. The need for more dependable critical infrastructures and better preparation has been emphasized in the task force reports of these and several other catastrophic events, specifically by tightening dependability requirements and incorporating cyber and socio-technical aspects in investigations.

In order to improve the dependability of CPSs, accurate models are needed. In a disruption cycle of a system, dependability models can substantially help at three stages:

1. Before disruption: Models help to raise awareness and understanding of the potential risks and their consequences, compare alternative recovery strategies, and prepare for contingency planning.

2. During disruption: Availability of knowledge and resources enable use of appropriate decisions to mitigate consequences and support rapid recovery.

3. After disruption: Models can help in determining high-priority actions required for restoration of essential services and resources needed for supporting recovery. Models should be continually refined based on what learned from the event.

Model-based analysis is a common and effective method for investigating failure scenarios of a system and evaluating its dependability attributes. Models can facilitate the comparison of alternative designs and expedite the design process; however, developing unified models of CPSs is a challenging task, as the model has to reflect hardware and software operation, as well as the continuous dynamics of physical systems [7]. Despite increasing activity in research related to CPSs, such models are still scarce, and to a large extent qualitative.

The overarching objective of this research is analytical modeling of dependability attributes of CPSs with domain-specific case studies on smart grids and intelligent water distribution networks. While both qualitative and quantitative models are important and useful for analysis of CPSs, this work focuses on quantitative modeling, as it has not been properly addressed in the literature. We evaluate the potential vulnerabilities and quantify the loss of dependability as a result of introducing computational and communication technologies . Furthermore, in order to compose a unified dependability model from the quantified dependability attributes, we study interdependence among the components of CPSs and investigate its effect on fault propagation paths.

Dependability is an integrative concept with multiple attributes, from which we seek to address reliability and survivability, as they have higher priority and are more appropriate in analysis of power and water critical infrastructures. Interdependency is the concept that attaches the disparate parts of dependability and provides integrity in model composition. Our contribution to the area of CPS dependability analysis is shown in Figure 1.2 and further explained in the following list. Respective publications are mentioned under each item.

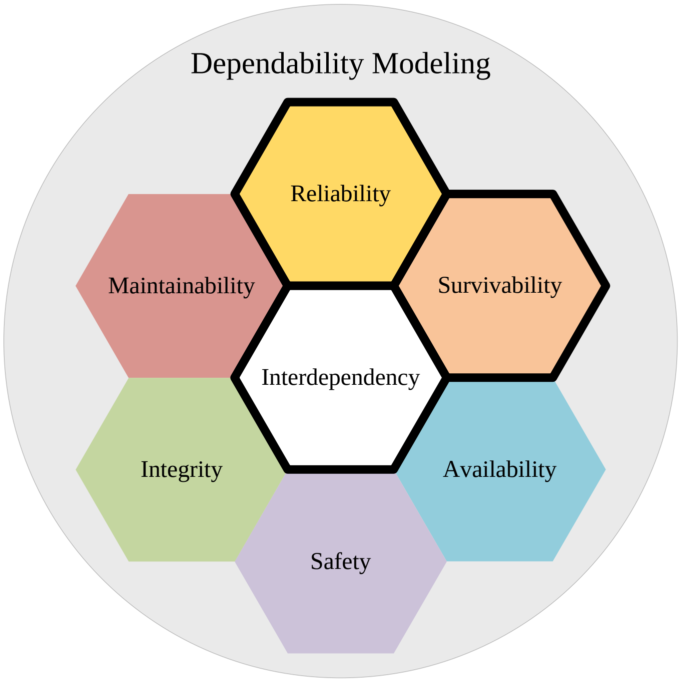


Figure 1.2. Our contribution to the area of CPS dependability modeling. Attributes that are covered in this research are shown in hexagons with thick borders.

1. Development of a quantitative reliability model using a Markovian technique that is applicable to CPSs with interdependent components

    - K. Marashi and S. Sedigh Sarvestani, "Towards comprehensive modeling of reliability for smart grids: Requirements and challenges," in *Proceedings of the 15th IEEE International High Assurance Systems Engineering Symposium (HASE)*, (Miami, FL), pp. 105–112, January 2014

  - K. Marashi, M. Woodard, S. Sedigh Sarvestani, and A. R. Hurson, "Quantitative reliability analysis for intelligent water distribution networks," in *Proceedings of the Embedded Topical Meeting on Risk Management for Complex Socio-Technical Systems (RM4CSS), Annual Meeting of the American Nuclear Society*, (Washington, D.C.), November 2013
  - K. Marashi, M. Woodard, S. Sedigh Sarvestani, and A. R. Hurson, "Quantitative reliability analysis for intelligent water distribution networks," in *Risk Management for Complex Socio-Technical Systems*, American Nuclear Society, to appear
  - K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, "Consideration of cyber-physical interdependencies in reliability modeling of smart grids," *IEEE Transactions on Sustainable Computing – Special Issue on Sustainable Cyber-Physical Systems*, to appear

2. Evaluating quantitative survivability attributes from service indices of CPSs

   - M. Woodard, K. Marashi, and S. Sedigh Sarvestani, "Survivability evaluation and importance analysis for complex networked systems," *IEEE Transactions on Network Science and Engineering*, under review

3. Proposing two methods using correlation metrics and causation analysis for identification of interdependency among components of a CPS

   - K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, "Identification of interdependencies and prediction of fault propagation for cyber-physical systems," *Reliability Engineering & System Safety*, to be submitted

4. Introducing interdependency metrics for quantifying the extent to which components and subsystems of a CPS are interdependent

- K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, “Quantification and analysis of interdependency in cyber-physical systems,” in *Proceedings of of the 3rd International Workshop on Reliability and Security Aspects for Critical Infrastructure (ReSA4CI 2016), in conjunction with the 46th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2016)*, (Toulouse, France), pp. 149–154, June 2016

5. Proposing a method using machine learning tools for prediction of failure sequences in interdependent CPSs

    - K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, “Identification of interdependencies and prediction of fault propagation for cyber-physical systems,” *Reliability Engineering & System Safety*, to be submitted

As well as the listed publications, we compiled a survey article on the recent research on modeling of CPSs as shown below.

- N. Jarus, M. Woodard, K. Marashi, A. Faza, J. Lin, P. Maheshwari, and S. Sedigh Sarvestani, “Survey on modeling and design of cyber-physical systems,” *ACM Transactions on Cyber-Physical Systems*, under review

Overall, as a result of this research three peer-reviewed conference papers [9, 8, 14], one book chapter [10], and four journal papers [11, 12, 15, 13] were published.

As mentioned, for demonstration of our proposed modeling approaches we will perform case studies on smart grid and intelligent water distribution networks. The term *smart grid* describes a modernized electrical grid that uses information and communications technologies and computer-based remote control to improve the efficiency, reliability, and sustainability of the production and distribution of electricity [16, 1]. Smart grids present an emerging solution to problems caused by increasing electric power demand from aging traditional power grids. Water distribution networks are constituents of another critical infrastructure that is considered an essential requirement of a modern city and a measure of the standard of living of the community. The primary goal of the water distribution networks is to provide a dependable source of potable water to the public. An *intelligent water distribution network* is the modern counterpart of the traditional water networks that collects and utilizes information on demand patterns, water quantity, and water quality in order to improve dependability, guide maintenance efforts, and identify vulnerable areas requiring fortification and/or monitoring. Despite difference in the physical commodities, smart grids and intelligent water distribution networks are common in structural and topological features, which enables the use of the same modeling approaches for both of these critical infrastructure CPSs.

The basis for our proposed methods is observation of system's behavior in a set of specified states. Given the high dependability expected of critical infrastructures and their large scale, fault injection studies on actual systems are infeasible, and organic failures are few and far between. As such, we use simulation platforms for determining the behavior of water and power systems in the presence of disruptions. Typical domain-specific simulation environments for power and water systems are incapable of capturing the behavior of the cyber infrastructure with the resolution required for analysis of CPSs. Therefore, we integrated MATLAB with these domain-specific simulators, namely, EPANET [17] for water distribution networks, and PSAT [18] for power grid systems, to enable holistic cyber-physical simulation.

Simulation environments enable us to inject hardware and software faults and study the behavior of the system in response to those disruptions. For each dependability attribute, associated behaviors are captured and populated into a model. Figure 1.3 shows the scope of our research and outlines our approach for modeling dependability aspects of CPSs.

In our reliability modeling [9, 10, 11], the set of plausible states are investigated and it is determined through simulation whether each state leads to an operational or failed system-level state. The results are then populated into a quantitative reliability model

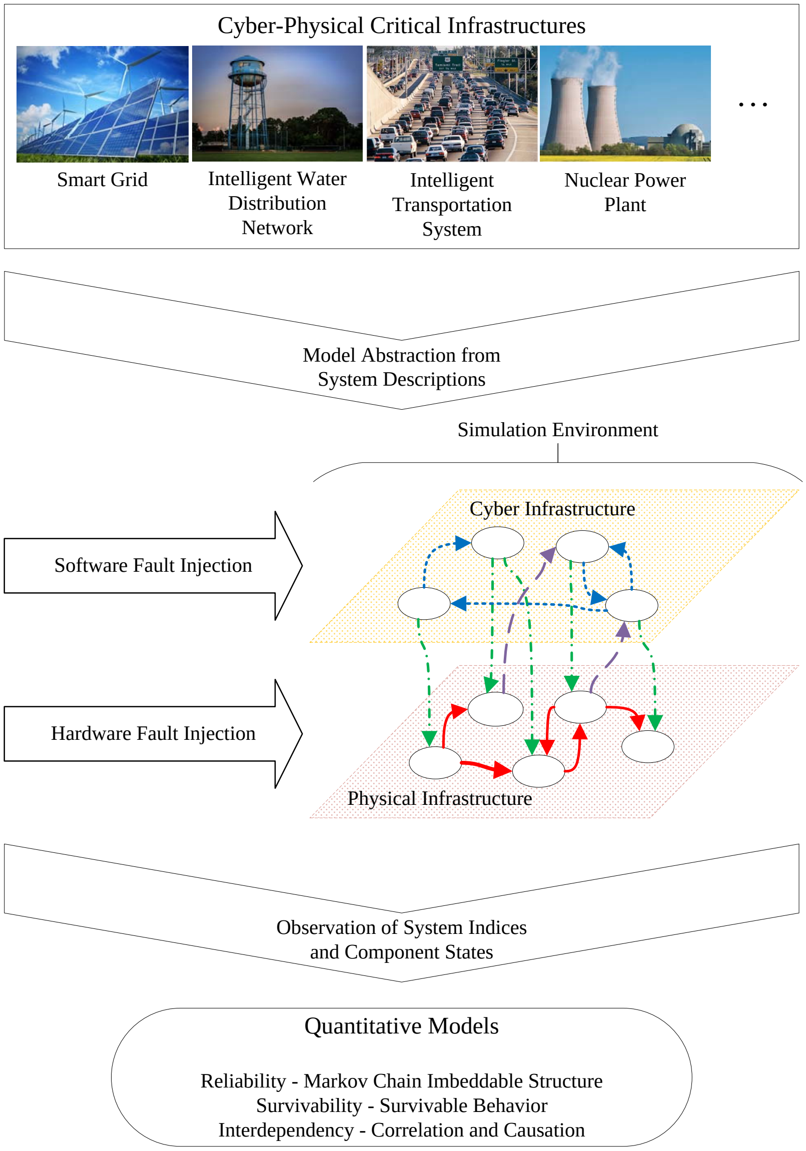

Cyber-Physical Critical Infrastructures
Smart Grid
Intelligent Water Distribution Network
Intelligent Transportation System
Nuclear Power Plant
...
Model Abstraction from System Descriptions
Simulation Environment
Cyber Infrastructure
Software Fault Injection
Hardware Fault Injection
Physical Infrastructure
Observation of System Indices and Component States
Quantitative Models
Reliability - Markov Chain Imbeddable Structure
Survivability - Survivable Behavior
Interdependency - Correlation and Causation

Figure 1.3. Our approach for dependability modeling of cyber-physical critical infrastructures.

using Markov chain imbeddable structure technique [19]. Unlike reliability modeling, our survivability evaluation technique tends to capture temporal behavior of the system and its capabilities in degraded states [12]. We quantify survivable behavior in terms of the *extent* and *rate* of degradation of a measure showing the level to which essential services are provided. Subsequently, components whose failure is the most detrimental to survivability are identified and fortified.

Similarly in our interdependency modeling [13], we observe sequence of events that occur after injecting software and hardware faults and find fault propagation patterns using correlation and causation analyses. We then quantify the interdependency using graph-theoretical metrics [14]. With the interdependencies revealed, we are able to take advantage of our dependability models in i) identifying weaknesses ii) determining how they can lead to vulnerabilities and propagate to other sections, iii) being prepared for potential risks, iv) making timely actions to mitigate the consequences of accidental failures and malicious attacks, and v) devising effective recovery strategies. We investigate the use of machine learning techniques in building a failure prediction tool, which can help system operators to better understand the risks and perform timely preventive actions [13]. Our approach is based on an artificial neural network trained with data collected from observed sequences of failure and is able to predict imminent failures given the current state of the system.

To illustrate the application of our proposed approaches, we have performed case studies on power and water systems. For the power domain, we chose well-studied IEEE 14-bus and 57-bus systems and supplied them with overlaid cyber infrastructure. Similarly for the water domain, we selected a test system studied in the literature and supplied it with intelligent control. Resulting CPSs do not include all of the existing cyber technologies used in respective areas, but are representative of real-world systems and are adequate for demonstration purposes.

What makes this research original and different from existing studies is consideration of cyber infrastructure as an error-prone entity with complex functionality, instead of reducing its role to a simple protective device. The quantitative nature of our models and the use of simulator-generated data, which alleviates the burden of procuring field data, are other distinctive features of this work.

The intellectual merit of this research is advancing knowledge of the potential loss of dependability due to the impairments originating in cyber and physical components, propagated through unprotected channels, and escalated into catastrophes in interdependent critical infrastructures. The broader impact of this research is a reduction in service interruptions and increased tolerance against disruptions for critical infrastructure CPSs. This is attained by providing models that enable engineers to better understand the consequences of their design decisions.

The remainder of this document is organized as follows. Section 2 discusses dependability attributes, summarizes related literature on modeling of CPS interdependency, reliability, and survivability, and articulates the uniqueness of our research in the context of the existing studies. Section 3 presents the methodology used in developing our models and Section 4 presents the results obtained from application of the proposed approaches to power and water systems. Section 5 outlines future research directions and concludes this dissertation.

2. BACKGROUND AND RELATED WORK

Dependability is a non-functional attribute that captures the behavior of a system during its life cycle. In the literature, dependability is defined as “the ability of a system to avoid service failures that are more frequent and more severe than is acceptable” [20]. It is an integrative concept that encompasses reliability, availability, safety, integrity, and maintainability. Each of these constituent attributes has a different definition and is meaningful for specific applications. For example, availability is an appropriate metric where a short outage of the service is negligible, but the system should be functional most of the time, e.g., web services. On the other hand, reliability is a meaningful metric where intermittent operation is not acceptable and the system should be continually functional, e.g., a power delivery system.

As indicated in [21, 20], survivability and dependability are concepts that are essentially equivalent in their goals and address similar threats with minor differences, thus survivability is considered as a dependability-related attribute.

Reliability and availability, two important aspects of dependability, consider the system state to be binary – *operational* or *failed*. This view is sometimes inadequate for large-scale CPSs such as critical infrastructures, which are expected to deliver uninterrupted service despite continual disturbances. It is expected that a large-scale system will spend time in functionally degraded states, without interruption of essential services. Consequently, additional non-functional attributes are required to characterize these degraded, yet operational states.

Performability, introduced by Meyer [22], combines performance and availability to evaluate system effectiveness, taking into account behavior due to failures. In other words, a system can be in a fully functional state, a partially operational state with degraded performance, or a failed state. Performability evaluates the expected performance over a duration composed of alternating operational/degraded/failed periods. Introducing performability as a metric that combines availability (pessimistic, binary view of operation) and performance (optimistic, neglects periods of inoperability) is an effort to attain a realistic view of the system. *Survivability* is another non-functional attribute that was introduced with a similar objective, and is used to characterize degraded operation. Survivability can be used to describe degraded operation at any point after a disturbance occurs, regardless of whether the disturbance is a fault tolerated by the system, or a failure that actually causes degradation. Another attribute closely related to survivability is *Resilience*, which is defined as the ability of a system to bounce back from failure [23, 24], but its application is limited to the recovery phase that follows a failure, not any period beforehand.

Among the dependability attributes, our focus is on reliability and survivability. We have analyzed reliability as it is a meaningful and commonly used measure for evaluation of the domains of our interest, i.e., the electric power delivery and water distribution networks. Besides the features captured by reliability, critical infrastructure CPSs are expected to autonomously defend against attacks, remediate the consequences of failure, and recover in a timely manner. Classical dependability attributes such as reliability provide coarse-grained characterization of these qualities, unlike survivability, whose very purpose is to precisely characterize transient behavior after a disturbance. For this exact reason, it has been used in several different domains including weapons systems engineering [25], telecommunication services [26], information systems [21], and software engineering [27].

Reliability has a formal definition from which a quantitative definition was derived. Reliability is a measure of the continuous delivery of correct service, and is formally defined as “the ability of a system or component to perform its required functions under stated conditions for a specified period of time” [28]. On the other hand, no standard definition of survivability was identified at the time of writing this dissertation; perspectives on the topic are diverse [29]. A concise qualitative definition presented by Heegaard and

Trivedi [30] states that "Survivability is the system's ability to continuously deliver services in compliance with the given requirements in the presence of failures and other undesired events."

Survivability has been quantitatively defined for networked systems by the ANSI T1A1.2 working group [31], using a domain-specific figure-of-merit (FoM), as shown in Figure 2.1:

> "Suppose a measure of interest M has the value m_0 just before a failure occurs. The survivability behavior can be depicted by the following attributes: m_a is the value of M just after the failure occurs; m_u is the maximum difference between the value of M and m_a after the failure; m_r is the restored value of M after some time t_r ; and t_R is the time for the system to restore the value of m_0."

To better illustrate the differences between the aforementioned dependability-related attributes, one operation and recovery cycle of a repairable system is shown in Figure 2.1. Note that $M(t)$ denotes an FoM chosen to represent the behavior or performance of the system over time. According to the definitions provided, the scopes of reliability and survivability, as well as other discussed attributes are indicated on Figure 2.1.

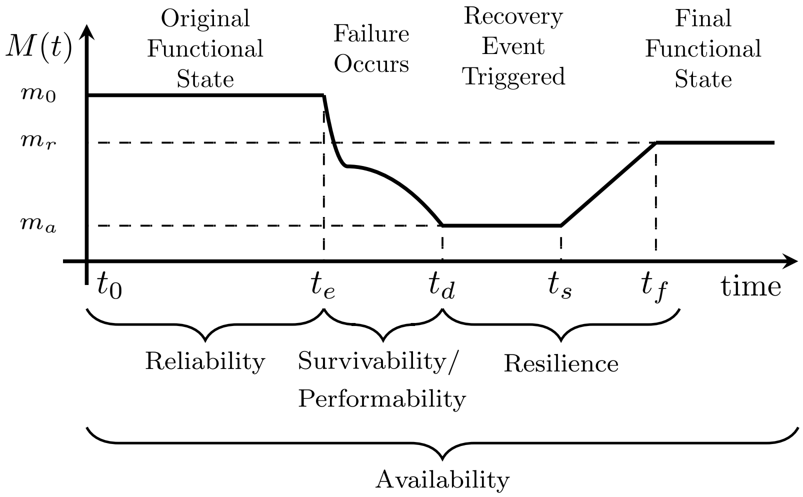


Figure 2.1. Durations of applicability for dependability-related metrics.

In the remainder of this section, the studies closest to the scope of our research are reviewed. Among the large amount of publications in the areas of our interest, we mostly focused on those that have been applied or are applicable to critical infrastructure CPSs, specifically the smart power grid. The related literature is organized as shown in Figure 2.2.

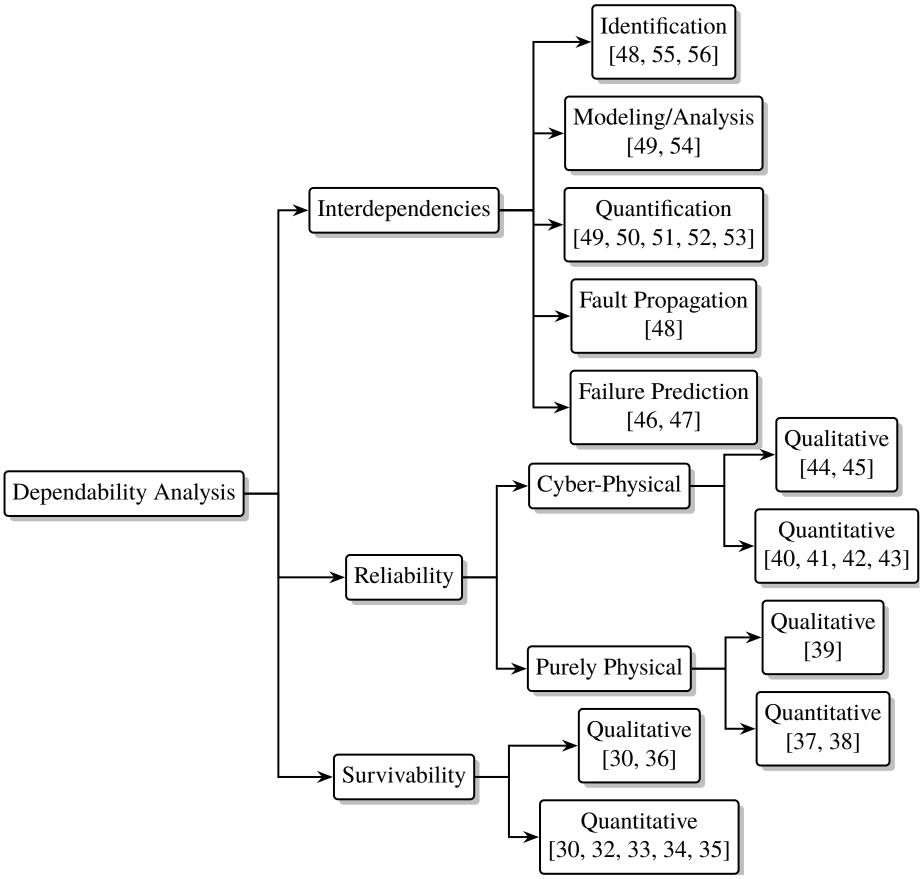


Figure 2.2. Organization of the related literature along with examples of each category.

2.1. RELATED WORK ON CPS INTERDEPENDENCY

There is an extensive literature devoted to the analysis of interdependency and its effects. The interdependence can be viewed as the relationships among the components of a single system [57], or among services provided by different systems [48, 58]. In the area of critical infrastructure, Rinaldi et al. have provided an ontology for understanding interdependence and classified its types into physical, cyber, geographic, and logical [59]. Similar classifications are presented in other studies, as enumerated in [60].

Researchers have used different approaches to identify interdependence among components, systems, or operations. In [48], the interdependencies between electrical infrastructure and the associated information infrastructure are qualitatively investigated and the pattern of fault propagation is explored. There are also examples of using correlation metrics for studying the interdependence. In [55], Pearson's correlation metric is used to investigate dependence among critical infrastructures after the World Trade Center attack. In another study [56], the time-series analysis is utilized to reveal interdependencies across critical infrastructures from post-event restoration curves of February 2010 Chile earthquake.

Models of interdependency are presented using a variety of techniques such as topology-based and flow-based methods, Bayesian networks, and Petri nets. In [49], Beccuti et al. proposed an approach using stochastic Petri nets for modeling the operation of the physical and cyber networks in electric power delivery systems. They subsequently measured the effect of disruptions of one network on the other (e.g., as a result of a cyber attack) in terms of a number of domain-specific performance indices. The study presented in [54] is an example of the application of dynamic Bayesian networks to analyzing interdependencies of critical infrastructures.

Another group of studies are devoted to quantification of interdependencies. Casalicchio and Galli have presented a number of quantitative metrics for interdependencies in critical infrastructures [50]. Studies presented in [51, 52, 53] use topological metrics (e.g., connectivity and size of giant component) to quantify interdependency in a network; however, for the case of power grids, Verma et al. argue that topological measures that are not context-aware may underestimate vulnerability of the system [61]. Additionally in [62], a comparison between blackout size and topological measures was performed for several power grid cases and only a mild correlation was observed.

As mentioned earlier, an important result of interdependency analysis is to predict the risk of failures for components and systems, prioritize preventive maintenance, and perform timely actions to mitigate effects of disruptions. In [46], statistical machine learning techniques were used to predict failure of feeder lines in boroughs of New York City over a three-month period and showed an acceptable accuracy of 75%. Additional results on application of the proposed methods are presented in [47].

2.2. RELATED WORK ON CPS RELIABILITY

Many researchers have studied the reliability of purely physical systems. It is important to clarify that there are abundant studies on analysis of critical infrastructures providing measures of reliability, however, they do not present quantitative models based on the formal definitions. Focus of this section is on reviewing the studies that present models only.

One classical analysis approach is the employment of Monte Carlo simulations. An example is [63], where an analytical model populated using a Monte Carlo simulation is presented for evaluating the reliability indices of power distribution systems. Fault tree analysis, another popular tool, is used in [37] for reliability evaluation of power systems. Bayesian networks approach is also a useful probabilistic tool for system reliability assessment that offers a transparent modeling scheme. It is used in [64], where a framework for system reliability assessment is presented and an algorithm is introduced in order to address the issue of exponential growth of the system states.

Graph-theoretical approaches have also been of interest to researchers for many modeling problems, as in [38], where authors have modeled the reliability of power grids. Their work aims at finding the most vulnerable nodes and edges with respect to attacks and accidental failures. They have formulated the reliability model considering both electrical indices (impedance of transmission lines) and reliability indices (probability of failure in network components).

The aforementioned studies do not examine the role of the cyber network. The reliability analysis techniques utilized in these studies are only applicable to purely physical system and may not be directly used for CPSs, which have heterogeneity and cyber-physical interdependencies [65].

In the last decade, extensive studies have been performed to determine the impacts of communication, information network, and computers on critical infrastructure CPSs. In [45], the author outlines at a very high-level, a mathematical framework for quantitative reliability analysis of cyber-physical power grids and compares potential approaches for accurate reliability modeling. The study presented in [66] discusses some of the complications due to the addition of the cyber layer to the power systems and investigates possibilities for transferring and utilizing the high volume of data gathered from numerous measurement devices effectively. Another study presented in [40] uses a pseudo-sequential Monte Carlo simulation and provides a tool for reliability assessment of smart power distribution systems considering failure of the communication infrastructure and short-term variations in distributed generation sites. An example of applying graph-theoretical methods to cyber-physical environments is [67], where the best topological configuration in terms of system reliability is investigated.

Generally, due to the diversity of the technologies that are nowadays being incorporated in a CPS, many of the recent reliability analyses focus on specific subsystems, e.g., the measurement subsystem [68] and communications [69].

In contrast, a number of studies have taken a holistic view of reliability for CPSs, taking consequences of disruptions in the cyber network into account to varying extents. For instance, [41] combines fault tree analysis, partial state space evaluation, and simulations to propose a quantitative reliability model for smart grids. The study focuses on reducing computational intensity and complexity of mathematical calculations by state merging and eliminating the states representing rare contingencies. This work however, is in an embryonic state and no tangible quantitative result is presented on a specific example.

Two other closely related studies are [42] and [43] that quantitatively evaluate the reliability of smart grids considering respectively direct and indirect interdependencies between power elements and cyber control devices. Shortcomings of [42, 43] are the consideration of limited and simplistic functions for the cyber network and overlooking interdependencies with intermediary events.

2.3. RELATED WORK ON CPS SURVIVABILITY

Comparison of survivability evaluation techniques is complicated by the lack of a common definition for this attribute. Knight et al. [36] has presented a survivability definition and model based on a service state-transition graph constructed from a six-tuple of service specification levels, service value factors, reachable environmental states, relative service values, set of valid transitions, and service probabilities. Ma [32] similarly quantifies survivability as a four-tuple of resistance, resilience, persistence, and failure count. In [32], *resistance* refers to the ability to withstand an attack, *resilience* refers to the mean recovery time from a catastrophic failure, *persistence* is the ability to maintain or exceed the minimal threshold of required functionality, and *failure count* describes the number of failures encountered over the duration of observation. In both of these studies the individual attributes are well-defined, but disjoint, and as such, none of them lead to a practical approach for quantitative evaluation of survivability.

Menasché et al. [33] have proposed an enhancement to the well-known SAIDI[1] metric. Their measure is denoted as *ESAIDI* and applied to evaluation of survivability aspects of a smart grid. ESAIDI is intended to facilitate analysis of the consequences of failures in distribution automation in the power grid. Avritzer et al. [70] utilize the same approach and further extend the model to account for disruptions in the communication infrastructure. This work was later combined with power flow analysis to create a survivability model that facilitates optimal design for the automation system of a smart grid [71]. A subsequent

[1] System average interruption duration index

extension of the work [72] allowed for concurrent failures in the power system. All of these approaches [33, 70, 71, 72] use time-to-recovery as a measure of survivability; however, time-to-recovery spans both the failure and recovery processes, and as such, cannot be used to separately evaluate system during each of these two phases.

Alobaidi et al. [35] evaluate the survivability of smart grids by studying the relationship between system condition (in terms of the number of functional components) and system capacity (ability to provide power to customers). The authors also propose and demonstrate recovery strategies intended to maximize survivability. The limitation of the proposed approach is that it is applicable only to power systems. In contrast, the work we present can be applied to any networked system with a known topology.

Chopade and Bikdash [34] present a model for survivability of a smart grid, based on graph-theoretic measures such as degree distribution and clustering coefficient. All buses (vertices) and lines (edges) are assumed to be identical. This is an unrealistic assumption given that reliability and other attributes of buses and lines can vary significantly in a power grid. In a similar vein, [73] evaluates the survivability of mobile ad hoc networks (MANETs) as the probability that all active nodes are k-connected to the network. This probability is determined using a semi-Markov model that captures state transitions due to node failures and malicious attacks. The connectedness of MANETs is a representative measure of their functionality; thus, the probability of being k-connected can reflect survivability. The proposed method is ill-suited to evaluation of any system expected to provide services beyond connectivity.

Avritzer et al. [29] survey recent approaches to survivability evaluation of water, gas, and electricity infrastructures. Stochastic hybrid models such as fluid stochastic Petri nets, hybrid Petri nets, and piece-wise deterministic Markov processes, as well as graph-theoretic approaches, are among the methods described.

2.4. THE DISTINCTION OF THIS RESEARCH

In spite of existing literature on dependability modeling of critical infrastructure CPSs, there are critical gaps that we seek to fill in this research. The missing parts of the literature, which are addressed in this dissertation are (i) quantitative dependability modeling, and (ii) consideration of cyber-physical interdependencies and analyzing the potential risk of increasing the failure propagation channels by deploying computational and communication technologies. An important feature of this work is consideration of the role of cyber infrastructure as a decision-making entity and not simply tasked with monitoring and protection. Another distinction is generality of the proposed approaches, which makes them applicable to multiple critical infrastructure domains.

3. METHODOLOGY

The focus of this research is on developing quantitative system-level reliability and survivability models for CPSs that accurately reflect the operation (and failure) of intelligent decision support and control, the physical operation, and the interdependency between the two. This work presents integrated quantitative models that capture impairments in both physical and cyber infrastructures. We also present a scheme of fault propagation in CPSs to identify the components that need to be reinforced in order to impede a cascade of failures.

In the remainder of this section, our approaches for identification and quantification of interdependencies, as well as analyses on reliability and survivability are described. It is worth mentioning that in the following discussions, we make assumptions specific to the smart grid domain (e.g., regarding the components and their roles); however, the approach is applicable to other domains.

3.1. ANALYSIS OF INTERDEPENDENCIES

Dependency is a linkage between two components, through which the state of one component influences or is correlated to the state of the other. In this case, the relationship is usually unidirectional, i.e., component i depends on j, but j does not depend on i. As an example, correct operation of a software that determines control commands for actuators in a robotic system is contingent on correct data from a sensor that detects surrounding objects. In this example, the software depends on the sensor, but the sensor does not depend on the software, as it can continue its operation regardless of the state of the software.

CPSs typically have an interconnected topology in which a bidirectional relationship may exist between the states of any given pair of components. *Interdependency* is a bidirectional relationship between two components through which the state of each component influences or is correlated to the state of the other. In other words, component i depends on j through a number of links, and j likewise depends on i through other links. More generally, two components are known to be interdependent when each is dependent on the other.

3.1.1. Representation of Interdependencies. For representation of interdependencies in a CPS, we use *dependency graph* that is a two-level weighted directed graph in which an edge exists from node i to node j if and only if the state of component i impacts, in one time step, the state of component j. We select time step small enough that we can assume impairment of a component propagates to the others through direct links only, not through multiple intermediate links.

Depending on the source and destination of an edge, it can represent one of four types of dependency, namely, physical-physical, physical-cyber, cyber-physical, and cyber-cyber. Note that in this notation, $s_1 - s_2$ dependence represents a relation in which components of subsystem 1 (s_1) influence components of subsystem 2 (s_2). As an example, Figure 3.1 illustrates the dependency graph for a hypothetical CPS.

In Figure 3.1, the bottom plane encompasses components of the physical system and the top plane is representative of the cyber infrastructure that monitors and controls the underlying physical processes. The weights shown on the edges represent the extent of dependency, denoted as the *degree of influence* and are in the range of $[0, 1]$, where a 0 means that there is no functional influence from a component on another, and a 1 is the case where the state of a component causes maximal degradation to the state of another, i.e., makes it unable to operate.

For mathematical representation, we introduce *direct influence matrix*, denoted as $\mathbf{D} = \left[d_{ij}\right] \in [0, 1]^{n\times n}$, which is in fact the adjacency matrix of the dependency graph. d_{ij} represents the *degree of influence* that component i exerts on component j and n is the total number of components in the system. Note that the entries on the diagonal of $\mathbf{D}$ should always equal zero, as a faulty state of a component "propagates" to itself immediately, not after one time step.

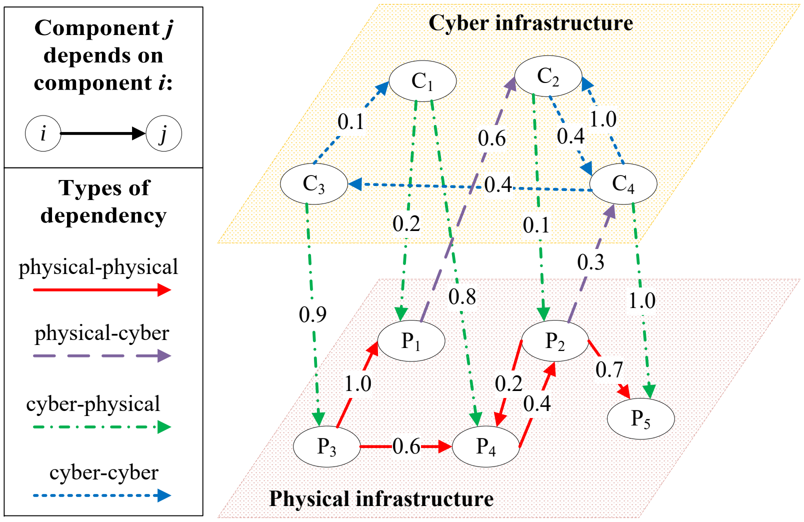


Figure 3.1. Dependency graph of a hypothetical CPS.

3.1.2. Identification of Interdependencies. Interdependency between components can be due to causality or simply a correlation. In a causation relationship, state of a component is responsible for that of another. On the other hand, state of two components are correlated when they have a statistical relationship, whether causal or not. Depending on the purpose of interdependency analysis, correlation or causation relationships may be of interest. In this section, we present two approaches for capturing correlation and causality, respectively, from observations of failure sequences corresponding to a set of failure cases.

A *failure case* is composed of a set of distinct components whose failure leads to disturbance to the system. These initial disruptions, exerted to the system at time t_e, may propagate to other components through dependency links, resulting in a *failure sequence*. Let $\mathcal{F}_k(t)$ represent the set of components that experience degradation during failure case k at time t.

In our approach, a set of failure cases are selected, and subsequent failures are observed. Dependency links are then extracted by analyzing these sequences. Note that a larger set of failure cases, and consequently, more observations will improve the accuracy of the interdependency model.

3.1.2.1. Correlation analysis. Statistical dependence between random variables that represent state of components is a potential tool for measuring interdependence. Conventional measures of dependence, such as Pearson's correlation coefficient, consider only a limited class of association patterns. More complicated correlation measures are able to detect non-linear relationships as well.

For each component we define a *state variable*, which is a random variable that characterizes the state of the component and subsequently, determine correlation between these random variables. Let $X_i(t)$ denote the state variable of component i at time t. For analysis of dependency of component j on component i, Pearson's correlation coefficient (PCC) between $X_i(t)$ and $X_j(t)$ is calculated, as shown in Equation (3.1).

$$r_{X_i X_j} = \frac{\mathrm{cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} \tag{3.1}$$

Where cov(.) is the covariance and σ is the standard deviation of a state variable. As explained in Section 3.1.1, we are interested in finding d_{ij} values, which represent dependence in one time step. Furthermore, direction of relationship (increasing or decreasing) is not of our interest. Therefore, we use $PCC_{X_i X_j} = |r_{X_i(t) X_j(t+1)}|$ to capture the direct dependency.

Shortcoming of PCC is that it only detects linear relationships, while an impaired component may result in disturbances, not necessarily linear, in another component. Among correlation coefficients introduced for detecting nonlinear relationships, we selected *randomized dependence coefficient* (RDC) [74], which has a low computational complexity and shows a good performance in comparison with similar methods. Readers are referred to [74] for more information on RDC. In this document, we use $RDC_{X_i X_j}$ notation to represent RDC correlation between state variables X_i and X_j. Note that RDC has two parameters associated with it, namely, sample size and number of random features, which can be set using guidelines provided in [74].

For identification of interdependencies we can compute mean value of correlation coefficients (either PCC or RDC) between X_i and X_j over all failure cases and use it as an estimator for d_{ij}.

3.1.2.2. Causation analysis. In general, a causal relationship is harder to establish than correlation, and hence, fewer interdependency studies have investigated causality. We use a method inspired by the interaction model introduced in [75], to identify causation relationships and estimate **D**. The work presented in [75], determines the interactions among components of a power grid, finds key dependency links, and provides strategies for mitigating cascading failures. In this section, we present a similar method that is generalized to be applicable for cyber-physical systems. Specifically, we have extended the method to incorporate heterogeneous components, control the sensitivity in detecting causality, and to account for dependency relationships between degraded states rather than binary states.

Consider a system composed of n components, for which m failure cases are observed. From the set of all failure sequences, i.e., $\mathcal{F}_k(t), \forall k$, we can construct matrix $\mathbf{W} = \left[w_{ij}\right] \in \mathbb{Z}^{n\times n}$, where w_{ij} shows the number of times component j has degraded one time step after degradation of component i over all failure cases. Components in the set $\mathcal{F}_k(t-1)$ whose states are known to be the dominant causes for degradation of component j in $\mathcal{F}_k(t)$ are identified using Equation (3.2).

$$\mathcal{H}_{k,j}(t) = \{\, i \mid i \in \mathcal{F}_k(t-1), w_{ij} \geq \eta \max_{l\in\mathcal{F}_k(t-1)} w_{lj} \,\} \tag{3.2}$$

In Equation (3.2), η controls the sensitivity in detecting the causative relationships, and is set to 0.9 in this work. Matrix $\mathbf{E} = \left[e_{ij}\right] \in \mathbb{Z}^{n\times n}$ is constructed as shown in Equation (3.3), where e_{ij} is the number of times degradation of component i caused degradation of component j.

$$\begin{aligned} \mathbf{E} &= \left[e_{ij}\right] \\ e_{ij} &= \sum_{k=1}^{m} \sum_{t>0} \operatorname{card}(\{\,(k,t) \mid i \in \mathcal{H}_{k,j}(t)\,\}) \end{aligned} \tag{3.3}$$

In Equation (3.3), card(.) denotes the cardinality of a set. Assuming that f_i is the total number of times component i experiences degradation over all failure cases, $\frac{e_{ij}}{f_i}$ estimates the likelihood of component i having a causative relationship with component j, i.e., the degree to which degradation of component i "causes" degradation in component j.

3.1.3. Quantification of Interdependencies. Assuming that the direct influence matrix $\mathbf{D}$ is known, we can explore dependency of components in multiple time steps. Specifically, we are interested in the *k^{th}-level influence matrix*, which represents the influence that components have on each other over exactly k time steps – in contrast to $\mathbf{D}$, where the influence exerted over a single time step is captured. To this end, we first normalize $\mathbf{D}$ by dividing it by n to ensure that $\sum_{j=1}^{n} d_{ij} \leq 1$. Matrix $\left(\frac{1}{n}\mathbf{D}\right)^k$ represents k^{th}-level influence matrix. The *total influence matrix*, $\mathbf{T}$, can be computed as shown in Equation (3.4).

$$
\begin{aligned}
\mathbf{T} &= \left[t_{ij}\right] = \mathbf{V} \circ \sum_{k=1}^{\infty} \left(\frac{1}{n}\mathbf{D}\right)^k \\
\mathbf{V} &= [v_{ij}] \\
v_{ij} &= \begin{cases} \dfrac{n+1}{n}, & i \neq j; \\ \dfrac{n+1}{n-1}, & i = j. \end{cases} , \quad 1 \leq i, j \leq n
\end{aligned}
\tag{3.4}
$$

In Equation (3.4), $\circ$ represents the entrywise product and t_{ij} shows the degree by which component j can be influenced by a failure in component i in any number of time steps, which reveals indirect influences. Note that the matrix $\mathbf{V}$ is used to scale t_{ij} to [0, 1] range.

In Equations (3.5) and (3.6), we define τ_i, and ν_j, which are respectively the weighted out-degree of node i and weighted in-degree of node j, in order to evaluate the extent of influence components exert on or receive from other components.

$$\tau_i = \frac{1}{n}\sum_{j=1}^{n} t_{ij} \tag{3.5}$$

$$\nu_j = \frac{1}{n}\sum_{i=1}^{n} t_{ij} \tag{3.6}$$

We will also measure the average dependence that components of subsystem s_1 have on components of subsystem s_2. For this purpose, $\gamma_{s_1-s_2}$ is calculated as shown in Equation (3.7).

$$\gamma_{s_1-s_2} = \frac{1}{n_{s_1} n_{s_2}} \sum_{i \in s_1} \sum_{j \in s_2} t_{ij} \tag{3.7}$$

In Equation (3.7), n_{s_1} and n_{s_2} are the number of elements in subsystems s_1 and s_2, respectively, and $n_{s_1} + n_{s_2} = n$. Note that τ, ν, and $\gamma_{s_1-s_2}$ are all normalized to the $[0, 1]$ range so that systems of different sizes can be easily compared.

3.2. RELIABILITY MODELING

A prerequisite for reliability modeling is to clearly define system-level failure. Assuming that such definition is provided, we can represent the system reliability as the probability of being in one of the system-level operational states. Our approach uses Markov chain imbeddable structure technique [19], which is an analytical method for reliability evaluation of systems with components whose reliabilities are known.

3.2.1. System State.

Ideally, all components of a system are in a fully functional state, however, this is not always the case. Each component may make a transition to another degraded state due to an internal or external disruption. Let C_i denote the set of states for component i with a cardinality of N_i. For a system composed of n components, *system state* can be represented by vector $\Phi = [\phi_i]$, where $\phi_i \in C_i$, $1 \leq i \leq n$. The system can be in one of $N = \prod_{i=1}^{n} N_i$ different states, hence, the system state space is represented by $\mathcal{S} = \{ \Phi_j \mid 1 \leq j \leq N \}$. Each of these N states can be classified as "operational" or "failed," as represented by the vector $\mathbf{U} = [u_j]$, where u_j is determined as shown in Equation (3.8).

$$u_j = \begin{cases} 1, & \Phi_j \text{ is an operational state;} \\ 0, & \text{otherwise.} \end{cases} \quad , \quad 1 \leq j \leq N \tag{3.8}$$

Assuming that $\Phi(t)$ is the system state at time t, the state distribution vector at time t is $\Pi(t) = [\pi_j(t)]$, where $\pi_j(t) = Pr\{\Phi(t) = \Phi_j\}, 1 \leq j \leq N$.

3.2.2. State Transitions. Starting from an initial state at $t = 0$, captured by $\Pi(0)$, the system can make transitions to other states with specific probabilities. These transition probabilities are in fact product terms composed of reliabilities and unreliabilities of the components. Due to the scope of reliability, state transitions are only considered until the system falls into a failed state, i.e., transitions in the repair phase are not incorporated. Let $\Lambda = [\lambda_{kl}]$ represent the transition probability matrix of the system, where $\lambda_{kl} = Pr\{\text{transition from state } k \text{ to state } l\}, 1 \leq k, l \leq N$. Similarly let Λ_i denote the state transition based on operation of component i. With this notation, there exist n transition probability matrices which collectively determine the ultimate state distribution of the system.

3.2.3. Markov Chain Imbeddable Structure. Assuming that the matrix $\mathbf{U}$ is known through investigation of all states in $\mathcal{S}$, the overall reliability of the system can be expressed as shown in Equation (3.9). It can be seen that R will in the form of sum of products, where each term represents the probability of being in one of the operational states.

$$R = \Pi(0) \left(\prod_{i=1}^{n} \Lambda_i \right) \mathbf{U}^T \tag{3.9}$$

Computational complexity for developing a model using Markov chain imbeddable structure technique is high; however, elimination of implausible states and use of data structures that fit this application (e.g., sparse matrices for Λ) will drastically mitigate the complexity. In a previous work [76], state aggregation as a complexity reduction method for reliability modeling has been investigated and found to be unnecessary for analysis of systems smaller than a specific size.

3.3. SURVIVABILITY MODELING

The high availability required in critical infrastructures makes it infeasible to bring down the system for fault injection studies. Detailed reports of real-world failures are few and far between, and many of the potential failure scenarios have never actually occurred in practice, necessitating the use of simulation tools. No simulation environment perfectly captures the characteristics of real-world entities; however, simulation does provide a good understanding of system behavior at minimal cost. This section introduces a survivability evaluation for CPSs, with an approach that can rely upon data from simulation, laboratory and/or field observation, and historical data about failures. We also present a method for identifying components whose failure is the most detrimental to survivability.

3.3.1. Survivability Attributes. Our survivability evaluation approach relies upon identification of a domain-specific FoM that is indicative of the extent to which one or more essential services are provided. In [31], *graceful degradation* and *failure resistance* are mentioned as two attributes essential to survivability. In defining metrics for survivability, we describe these attributes (with reference to Figure 2.1) as follows:

- *Graceful degradation* is achieved when the *rate of degradation*, $\left|\frac{dM(t)}{dt}\right|$, after a disturbance is considered to be slow, in the context of the time scale of the system domain.

- *Failure resistance* indicates that the *extent of degradation*, $|M(t_d) - M(t_e)|$, after a disturbance, i.e., the loss in FoM value incurred between the start of the disturbance and initiation of recovery, leaves the system functionality at an acceptable level.

The FoM is domain-specific, as it is intended to capture the extent to which a system is delivering essential services. In this work, we consider the FoM to represent a single service. Our survivability evaluation approach can be used to represent more complex behavior by defining an FoM that is a composite, e.g., a weighted average, of metrics that reflect different essential services.

3.3.2. Evaluation of Survivability. Graceful degradation and failure resistance are two attributes that are pivotal to our proposed approach to survivability evaluation and component importance analysis. We evaluate survivability through the following actions, carried out consecutively, seeking to quantify these attributes.

1. A system-specific FoM and a set of representative failure cases are selected to evaluate the system.
2. Each failure case is observed or simulated, and the value of the FoM is monitored over an interval that begins with a fully functional system, includes a disruption that causes degradation to the FoM, and continues through initiation of recovery efforts.
3. The rate and extent, respectively, of degradation of the FoM are calculated from the log of FoM values.

In each failure case, faults are injected to a set of distinct components at time t_e (as in Figure 2.1). Let $\mathcal{F}_k(t)$ represent the set of components failed at time t during failure case k. We consider component-level operation to be binary, i.e., a component is either fully functional or has failed altogether. This assumption is justified where the system representation is fine-grained and the contribution of a single component to delivery of an essential service cannot be further decomposed.

Exhaustive examination of failure cases is infeasible for large complex systems. On the other hand, omission of failure cases with catastrophic consequences could render survivability evaluation meaningless. This state space explosion problem is common in any type of system evaluation and its resolution is not within the scope of this dissertation. In this work we will assume that we have a predefined set of failure cases as the basis of survivability evaluation.

Suppose we have a system with n components and m failure cases that have been designated as the basis of survivability evaluation. Each failure case is observed or simulated for a duration that begins with a fully functional system where all components are operational, continues through the disturbance caused by failure of the components in $\mathcal{F}_k(t_e)$, and ends when recovery efforts are initiated. In other words, observation or simulation of the failure case k produces a record of the FoM, $M_k(t)$, for $t_0 \leq t \leq t_d$, where t_0 and t_d are as defined in Figure 2.1. It is worth noting that the failures of the components initiating the disturbance at t_e, $\mathcal{F}_k(t_e)$, can lead to failures of other components. This larger set of components, denoted as $\mathcal{G}_k$, includes any component whose failure is observed between t_e and t_d.

Survivability analysis requires that the $M_k(t)$ be examined to determine the extent and rate of degradation. In terms of Figure 2.1, we seek to determine the full extent of degradation, denoted as δ_k, incurred between the instant of disturbance (t_e) and initiation of recovery (t_d). Over the same period, the most rapid rate of degradation is denoted as ρ_k. Equations (3.10) and (3.11), respectively, reflect these attributes. The survivability of a system is determined by aggregating the extent and rate of degradation for all failure cases.

$$\delta_k = \max_{t_e \leq t \leq t_d} |M_k(t_0) - M_k(t)| \tag{3.10}$$

$$\rho_k = \max_{t_e \leq t \leq t_d} \left| \frac{dM_k(t)}{dt} \right| \tag{3.11}$$

Visualization of the FoM, as in Figure 3.2, facilitates evaluation of survivability. For each failure case, a degradation point, (ρ_k, δ_k), is used to calculate the *degradation index*, defined as the distance from the degradation point to the origin. The single degradation point (failure case) shown in Figure 3.2 has $\rho = 0.25$, $\delta = 0.6$, and a degradation index of 0.65. The degradation index facilitates comparison of failure cases and can be averaged across all failure cases to calculate a single survivability index for the system.

Creating a two-dimensional color intensity histogram of the degradation index, over all failure cases considered, can facilitate identification of clusters indicative of failure cases that are similar in consequence. In an ideal system, only one cluster would be evident, near

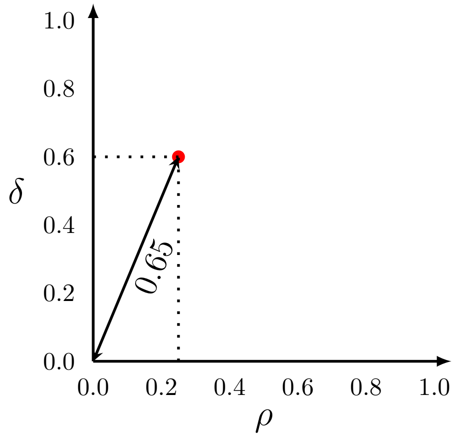


Figure 3.2. Rate and extent of FoM degradation, (ρ, δ), for a failure case.

the origin, in the lower left corner of the plot. This cluster is characterized by slow and minimal degradation of the system. Clusters outside of this area represent failure cases that merit further investigation, as they reflect non-survivable behavior.

3.3.3. Component Importance Analysis. Evaluation of survivability can illuminate weaknesses in a system. Specifically, our method can facilitate identification of components most in need of fortification, i.e., *importance analysis*, where the measure of importance is the contribution of a component to survivability. We propose two criteria for ranking components, namely, criticality and fragility.

The *criticality* of a component is determined by the consequences of its failure on service degradation, evaluated over all failure cases in which the component experiences failure. Recall that associated with each failure case k is a set, $\mathcal{G}_k$, that encompasses all components observed to fail during the failure case. As described in Section 3.3.2, the highest degradation incurred during a given failure case k is denoted as δ_k. To determine the criticality of component i, we need to identify every failure case in which it was observed to fail, the set of these cases, $\mathcal{Q}_i = \{ k \mid i \in \mathcal{G}_k, 1 \leq k \leq m \}$. Additionally, let $t_i^{(k)}$ denote the time at which component i has failed during failure case k. The criticality of a component is composed of three terms: The first term normalizes the extent of degradation to rank the severity of the failure case. The second term normalizes the rate of degradation at the instant of component i's failure during failure case k. The third term normalizes the second derivative of the FoM at the instant of component i's failure during failure case k. We consider this term to be indicative of the immediate impact of this specific component's failure during the failure case. The product is calculated and summed across all failure cases involving component i and divided by m, the total number of failure cases. The criticality of component i, α_i, is determined as shown in Equation (3.12).

$$\alpha_i = \frac{1}{m} \sum_{k \in Q_i} \left(\overbrace{\frac{\delta_k}{\max\limits_{1 \le l \le m} \delta_l}}^{\text{first term}} \cdot \overbrace{\frac{\left.\frac{dM_k(t)}{dt}\right|_{t=t_i^{(k)}}}{\max\limits_{\forall t} \frac{dM_k(t)}{dt}}}^{\text{second term}} \cdot \overbrace{\frac{\left.\frac{d^2M_k(t)}{dt^2}\right|_{t=t_i^{(k)}}}{\max\limits_{\forall t} \frac{d^2M_k(t)}{dt^2}}}^{\text{third term}} \right) \tag{3.12}$$

A less precise measure of the importance of a component is provided by *fragility*, which reflects the fraction of observed or simulated failure cases in which the component has failed. The fragility of component i, denoted as β_i , can be determined as shown in Equation (3.13), where m is the total number of failure cases.

$$\beta_i = \frac{|Q_i|}{m} \tag{3.13}$$

Either criticality or fragility can be used to determine the priority of a component for hardening efforts. Given that fragility is calculated without consideration of service degradation (as represented by the FoM), its use is recommended only in cases where failure information does not involve the exact time when each component failed.

3.4. PREDICTION OF FAILURE SEQUENCES

Upon availability of sufficient failure data and knowledge on interdependency among the components of a system, a prediction tool may be used to detect catastrophic failures in their incipient stage and enable the supervisory control team to perform timely preventive actions and make appropriate decisions to mitigate the consequences. For this purpose, powerful and reliable tools are needed that are capable of identifying the components (or sections of the system) that are prone to failure as a result of a disruptive event. Furthermore, such tools are expected to respond in real-time and provide a prioritization of the components that are in risk, based on their failure likelihood and importance of their roles in the system.

3.4.1. Preparation of the Failure Data. Assuming that failure data of a system is available, we need to convert them into a data set composed of several input/output entries to be used for training the predictor tool. Figure 3.3 shows an example of a failure sequence for a hypothetical system with eight components. In this failure case, the system initially has faults in components 2 and 6. The faulty state is propagated to other components and affects component 1, then components 4 and 8, and finally component 5. We transform this failure case into four entries of the data set used for training the predictor tool, as shown in the right of Figure 3.3. Each entry of the data set has two fields: i) an array of the state variables at a time instant and ii) a list of components that will degrade at the next time step.

3.4.2. Artificial Neural Networks Approach. The problem of predicting a sequence of events is closely related to classification and sequence labeling in time series analysis. In this work, we transform the problem of sequence prediction to a multi-class classification and investigate the use of artificial neural networks (ANN) for tackling this problem. Reports show that the ANN is a promised tool for classification problems [77]. In a multi-class classification problem, a given instance is to be associated with a number of classes. The classification can also be probabilistic, where the classifier provides

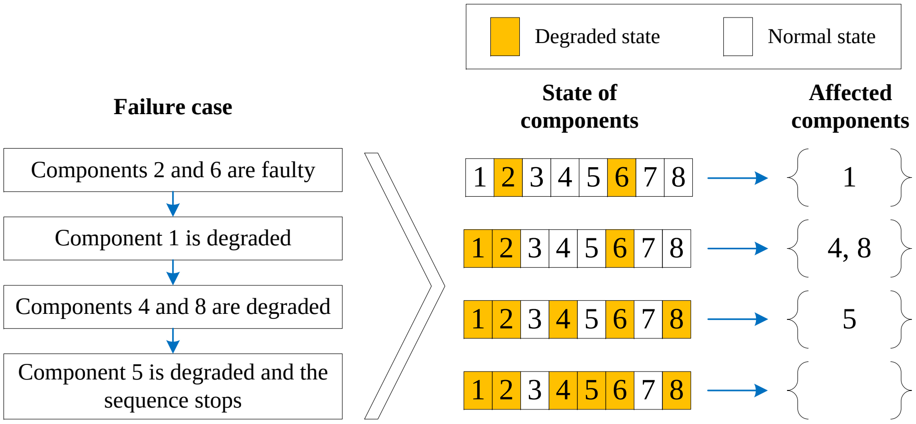


Figure 3.3. Transformation of a failure sequence into four entries of the data set used for training the failure prediction algorithm.

probability distribution of a given instance belonging to the existing classes. Some of the popular classification problems are speech recognition, pattern recognition, and several other applications in medical imaging.

For a system with n components, let $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_n(t))$ denote the input array to the ANN, where $X_i(t)$ is the state variable of component i at the time instant t. $\mathbf{X}(t)$ is fed to a multi-layer fully connected ANN with the architecture shown in Figure 3.4. The Output layer provides $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, where Y_i represents the probability that component i fails as a result of disruption specified by the given state variables in the input.

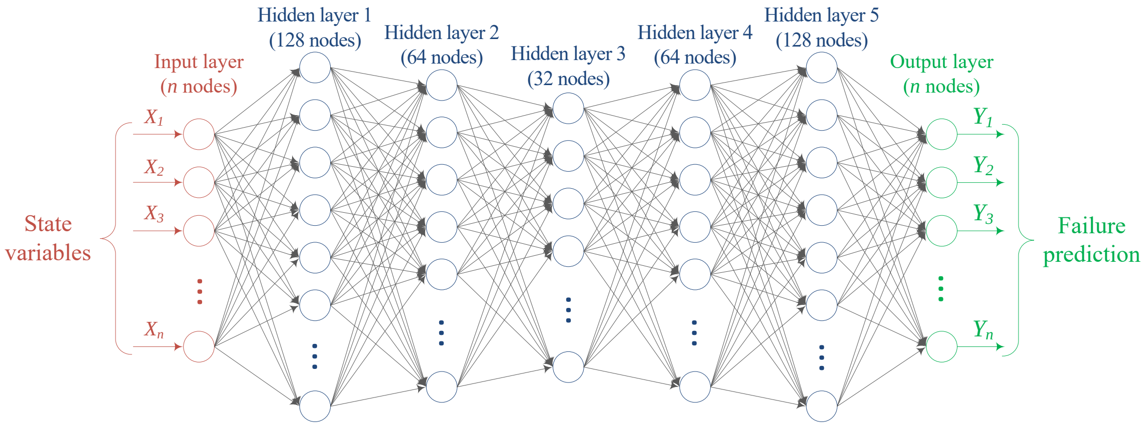


Figure 3.4. Architecture of the multi-layer fully connected ANN used for failure prediction.

In the nodes of the hidden layers, we have used the softplus activation function [78], which is a differentiable and smooth version of the well-known rectifier function, to introduce nonlinearity to the ANN. The optimizer used for updating weights of the ANN is Adam, as introduced in [79] and the loss function is found by calculating the cross entropy between sigmoid of the failure predictions and that of the actual failures. In order to prevent overfitting during the training process, we utilized L_2 regularization method, which penalizes the network for large weights by increasing the loss.

The choice of ANN architecture is generally based on heuristic rules and is only for the sake of demonstrating applicability of the method. The architecture described here has shown an excellent performance on the test cases investigated in Section 4.5; however, depending on the type and size of the system under test, adjustments may be required.

The ANN is trained using a data set generated by simulating a number of failure cases or data from historical information of previous disruptions. In either case, each entry of the data set should include state variable of the components at the time of disruptive event (input to the ANN), linked with the list of components affected consequently (used as ground truth for optimization during training and final verification).

Depending on the type of the system, preventive actions may be prioritized based on different parameters. Examples of prioritization parameters are the predicted failure probability of each component (provided by the neural network), importance of each component in providing essential services, and consequences of failure of each component on other components (e.g., weighted out-degree explained in Section 3.1.3) as well as on the operation of the system (e.g., in terms of loss of dependability [12]).

3.4.3. Evaluation of Predictive Performance. In order to evaluate the effectiveness of the proposed ANN in predicting failures, we should use metrics that capture the predictive performance. To this end, we take advantage of the available metrics in the areas of information retrieval and classification, namely, precision, recall, and F_1 score.

Precision shows the ratio of successful failure detections over the total detections. The precision has a shortcoming in evaluating the performance when the ANN correctly predicts only a portion of the failed components, but fails to detect the remainder. *Recall*, also known as sensitivity, is the ratio of the failed components detected by the ANN to the total number of failures. The shortcoming of recall is that its value is large if the ANN simply predicts that all of the components will fail. Therefore, no single metric is enough for evaluating the performance correctly. F_1 *score* has been introduced to solve this issue by combining precision and recall into a single metric by taking their harmonic mean. Equation (3.14) shows how these metrics are calculated.

$$\begin{aligned}\text{Precision} &= \frac{tp}{tp+fp}\\ \text{Recall} &= \frac{tp}{tp+fn}\\ F_1 \text{ score} &= 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\end{aligned} \tag{3.14}$$

In Equation (3.14), tp, tn, fp, and fn represent the numbers of true positives, true negatives, false positives, and false negatives, respectively. To better understand these terms, let us consider a commonly used example of binary classification. In this example the task of the classifier is to determine whether a prisoner is guilty or innocent. A *true positive* is the case where the prisoner is actually guilty and is correctly detected as guilty. If the guilty prisoner is freed from the jail, it is considered a *false negative*. On the other hand, convicting an innocent person is an example of *false positive* and freeing an innocent is a *true negative*.

4. RESULTS AND ANALYSIS

In this section, we first present and discuss the results obtained by applying our proposed approaches, described in Section 3, to smart power grid systems. We then explore applicability of the same approaches to other critical infrastructure CPSs, as a specific example, illustrate reliability modeling for an intelligent water distribution network.

4.1. CASE STUDY ON SMART GRIDS

For our case study, we have constructed two smart grids based on test systems well-studied in power engineering literature, namely, the IEEE 14- and IEEE 57-bus test systems [80]. To better understand the structure of a power delivery system and recognize threats to its dependability, we review common sources of failure in the domain of smart grids. The IEEE 14-bus system has been included in the interest of brevity and clarity and the IEEE 57-bus system demonstrates the scalability of our methods.

4.1.1. Potential Sources of Failure in Smart Grids. Rapid developments in generation and consumption of power are causing increasing stress on distribution networks. Among other benefits, cyber control brings more efficient use of the limited capacity available; however, each additional component used in this cyber control is a potential source of failure, and the net effect of this increased vulnerability and complexity on the overall reliability of the grid requires careful examination. A comprehensive dependability analysis should be able to consider every potential source of failure and reflect its effect on the overall system state. The remainder of this section enumerates the main categories of components that comprise a smart grid and can affect its operation by causing or increasing the likelihood of failure. The sections of a smart grid along with respective examples of potential failures are depicted in Figure 4.1.

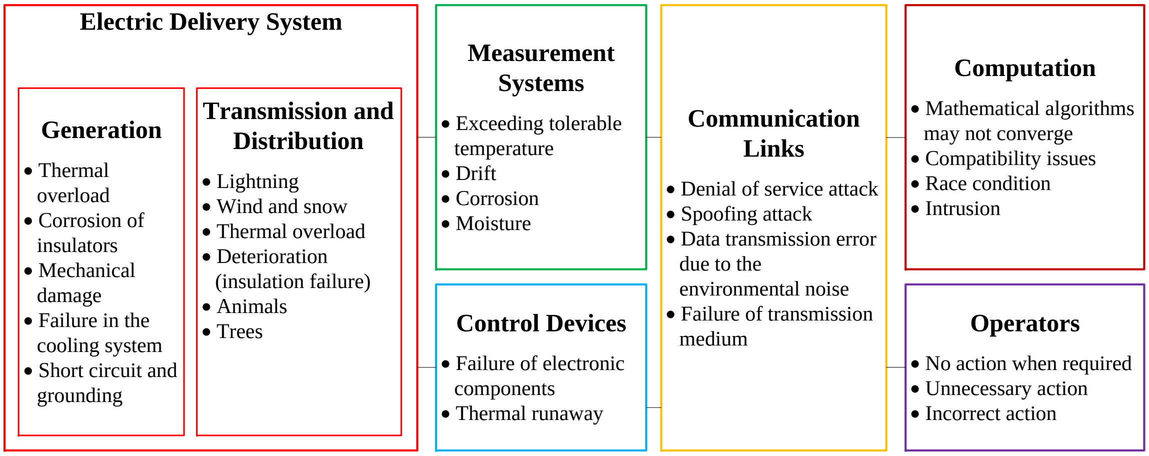

Electric Delivery System
Generation
• Thermal overload
• Corrosion of insulators
• Mechanical damage
• Failure in the cooling system
• Short circuit and grounding
Transmission and Distribution
• Lightning
• Wind and snow
• Thermal overload
• Deterioration (insulation failure)
• Animals
• Trees
Measurement Systems
• Exceeding tolerable temperature
• Drift
• Corrosion
• Moisture
Control Devices
• Failure of electronic components
• Thermal runaway
Communication Links
• Denial of service attack
• Spoofing attack
• Data transmission error due to the environmental noise
• Failure of transmission medium
Computation
• Mathematical algorithms may not converge
• Compatibility issues
• Race condition
• Intrusion
Operators
• No action when required
• Unnecessary action
• Incorrect action

Figure 4.1. Major sources of failure in a smart grid, by category.

4.1.1.1. Electrical infrastructure. Electric delivery systems are primarily composed of current-carrying components, including generators and transmission lines. In several studies, transmission lines are assumed to be the main sources of vulnerability, since generation units and similar components typically have enough backup to compensate for their failures [81, 82]. With this assumption, reliability analysis of a power grid usually entails tripping transmission lines, one-at-a-time, and inspecting the resulting state of the system. This process is also referred to as $N-1$ *contingency analysis*.

4.1.1.2. Control devices. Power flow control has traditionally relied on generator control and voltage regulation by means of phase-shifting transformers. These techniques are often found to be ineffective, as they rely on a permanently fixed configuration and lack the adaptability required in smart grids. Flexible AC transmission system (FACTS) is a technological development in electrical power systems that is based on the incorporation of power electronic devices for controlling the power system. Early developments of the FACTS technology were in power electronic versions of the phase-shifting transformers. Unified power flow controller, static compensator, and static synchronous series compensator are some examples of FACTS devices.

4.1.1.3. Communications. One important feature of smart grids is the integration of high-speed and reliable data communication networks to manage the complex power grid effectively and intelligently. The dependability of the power management is hence contingent on a reliable communication backbone. Various communications technologies are utilized at different sections of smart grids. Figure 4.2 depicts communication links in a conceptual smart grid, categorized into subsystems.

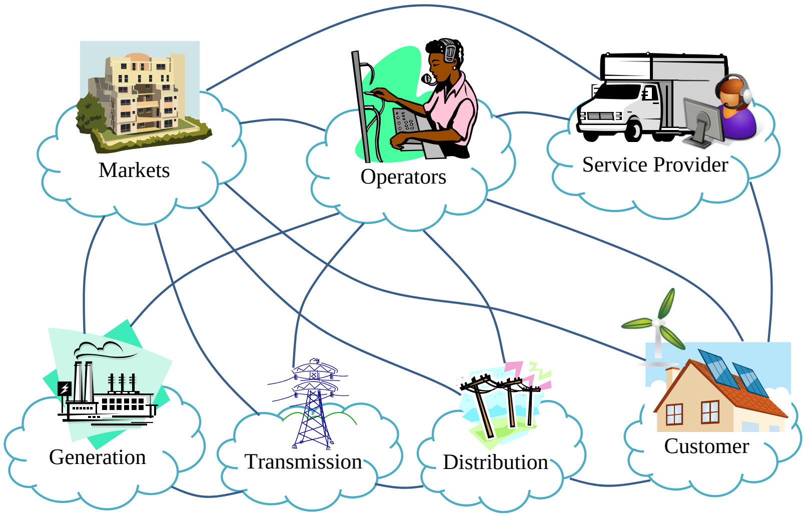


Figure 4.2. Communications in a smart grid (adapted from [1]).

In a smart grid, generation units communicate with the market and the operation domains. The information communicated includes key parameters such as generation capacity and shortage. In the transmission system, a significant amount of data is captured from the grid and sent to the control centers. The control centers in turn send responses to devices in remote substations. Distribution networks interact with multiple entities, such as distributed generators, automatic metering infrastructure, and sensors in order to enable provision of high-quality and stable electricity for the end-users. Operation domain mainly communicates to the transmission and distribution systems to obtain information about power system activities such as monitoring, maintenance, and metering. Market domain needs to communicate with the bulk producers of electricity and the distributed generating resources to match the production with the demand. The customer domain communicates with the distribution, operation, service provider, and market domains to facilitate remote load control, monitoring of distributed generators and in-home display support [68]. Finally, service providers communicate with the operation domain for situational awareness and system control and to obtain metering information. They also communicate with the customer domain to provide smart services such as management of energy use and home energy generation.

4.1.1.4. Measurement systems. Advanced sensing and measurement technologies evaluate the health of equipment and the integrity of the smart grid. The use of enhanced measurement and control allows the system to operate closer to its physical limits and increases its efficiency [45].

For better wide-area situational awareness, regional transmission operators require information about the state of the power grid, which is achieved by specialized sensors that provide real-time data - phasor measurement units (PMUs)- at substations. PMU devices capture current and voltage phasor information from the electrical buses at selected substations at sample rates of up to 60 Hz. The information received from PMUs is used by energy management systems at control centers for improved state estimation, monitoring, control, and protection.

Corruption of data received from the sensors or lack thereof, can result in erroneous situational awareness or unobservability to the decision support, resulting in inappropriate control of regulators and potentially a cascade of failures. Therefore, it is very important to assure that the measurement devices are highly reliable, calibrated, and maintained in specified intervals.

4.1.1.5. Computation. The evolution and use of decentralized control significantly complicates analysis of the large-scale distributed networks. It also necessitates that communication links and data transfer functions be considered alongside computing elements in dependability analysis. Software engineering has enabled the development of nearly-perfect computer programs that utilize control algorithms to optimize the functionality of a CPS; however, software faults often lead to system failures and should be considered in the dependability analysis.

4.1.1.6. Operators. If proper planning criteria are followed most modern power systems are designed to be able to operate safely and in a stable fashion with minor contingencies; however, depending on the severity of a failure event, the system may enter into an emergency state where a human operator needs to take an action. It should be noted that human errors are inevitable and can cause catastrophic failures in critical applications [83]. Therefore, dependability analysis should take into account the effect of human errors as well.

4.1.2. IEEE-14 and IEEE-57 Smart Grids. The IEEE 14-bus system consists of two generators supplying the grid with active power; 14 buses providing electricity for 11 loads; and 20 transmission lines interconnecting these buses. In order to study the effects of utilizing intelligent control systems and communication and information technology, we construct a smart grid from this physical system by incorporating cyber components. According to the method presented in [84], four PMUs are placed on buses 2, 4, 6, and 9. This placement strategy provides power system observability for the control systems and brings redundancy. Furthermore, based on the method presented in [85], three static synchronous series compensator (SSSC) devices are installed on specific transmission lines as depicted in Figure 4.3. SSSC is a type of FACTS device that is used for controlling the power flows in the network.

An ideal decision support algorithm utilizes the available resources and computing capabilities efficiently in order to mitigate faults and prevent cascading failures. Computational intelligence comprises promising approaches to solve the intricate problem of controlling the complex networks using the information collected from the sensors, considering the “global” effects of “local” decisions in an intelligent way [86]. In our case study,

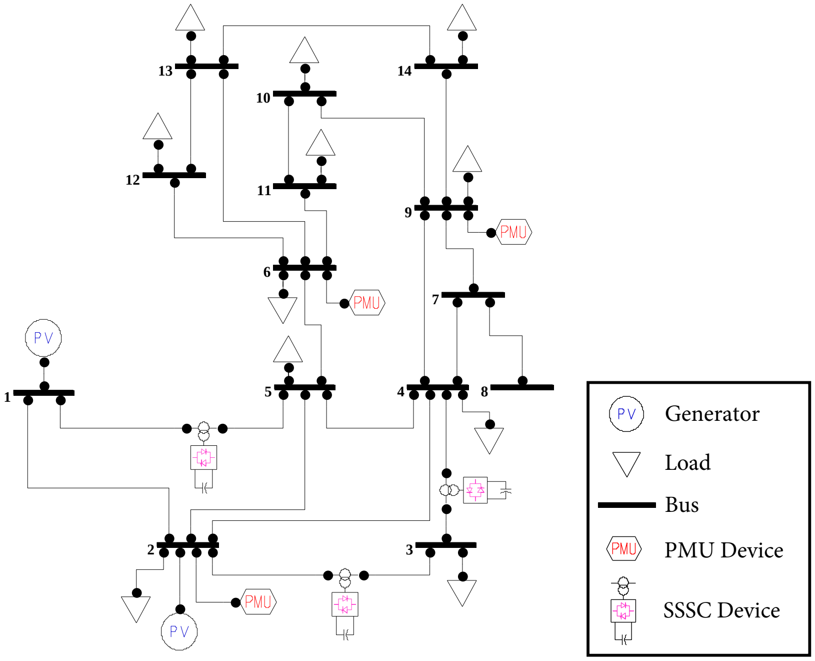


Figure 4.3. IEEE 14-bus smart grid.

an ANN trained with $N-1$ contingencies is used as the decision support algorithm. Finding the optimal setting for each SSSC device in all possible contingencies is a computationally intensive task. This ANN dynamically controls the SSSC devices according to the real-time measurements from PMUs with the objective of distributing the flow of power through all available routes and minimizing the risk of overloading. Applicability and performance of ANN in dynamic tuning of SSSC devices in unusual operating points are verified in [87]. Power utility companies mitigate the impacts of peak demand and impairments by load shedding, load balancing, and line current balancing. This ANN plays a similar role in our smart grid example and performs the same tasks autonomously. More details on the architecture of this ANN is provided in Appendix B.

Table 4.1 highlights the performance of the ANN in practice by comparing the number of cases that lead to system-level failure out of all simulated cases for three configurations, namely, IEEE-14 without any control devices, IEEE-14 with three fixed-tuned

SSSC devices installed on specific lines as shown in Figure 4.3, and lastly, IEEE-14 with the same three SSSC devices but dynamically tuned by the ANN according to the instantaneous system state.

Table 4.1. Number of cases out of all simulated cases that lead to system-level failure for IEEE-14.

	No SSSC devices	With fixed-tuned SSSC devices	With dynamically-tuned SSSC devices
Single-line outages that lead to a cascading failure	2 / 20	1 / 20	1 / 20
Double-line outages that lead to a cascading failure	44 / 190	29 / 190	23 / 190

It is seen in Table 4.1, that utilizing fixed-tuned SSSC devices can eliminate some of the vulnerabilities, as reflected by the reduction of the cases that lead to system-level failure, and hence, increases the robustness of the system against single- and double-line contingencies. Employing the ANN for dynamic tuning of SSSC devices further increases the robustness against double-line outages, as manifested by reduction of 29 cases leading to system-level failure down to 23.

The IEEE-57 is a larger power system that consists of seven generators; 57 buses providing electricity for 42 loads; and 80 transmission lines interconnecting these buses. Twelve PMUs and seven SSSC devices are also installed on specific buses and transmission lines according to the methods presented in [84] and [85], respectively, as illustrated in Figure 4.4. Similar to the IEEE-14 smart grid, an ANN is employed to serve as the decision support algorithm. The ANN utilized for the IEEE-57 smart grid however, has larger input, hidden, and output layers as it is responsible for controlling a larger system.

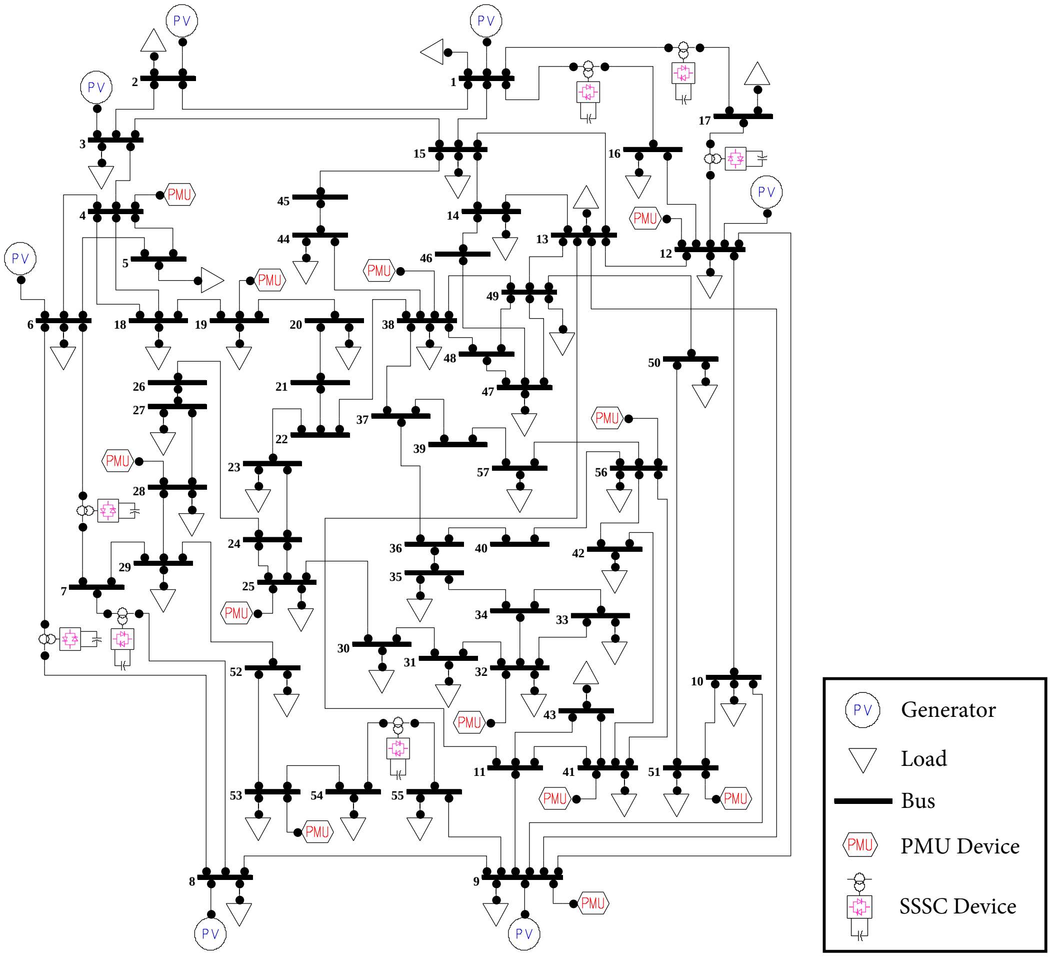


Figure 4.4. IEEE 57-bus smart grid.

In the remainder of this document, we use the following notations for the components of smart grid that are in the scope of our analysis.

- L_{i-j}: A transmission line connecting bus i to bus j
- F_{i-j}: An SSSC device installed on line L_{i-j}
- P_i: A PMU installed at bus i
- CM: A communication link between two cyber entities
- DS: The decision support

4.1.3. Integrated Cyber-Physical Simulator. A CPS simulator should be able to reflect the operation of multiple aspects: sensors, data collectors, computers, control systems, databases, communications, and the processes in the underlying physical infrastructure. Differences between cyber and physical components complicate representation of their behavior with a single simulation tool. For most sectors of the critical infrastructure, including power, water, and transportation, specialized simulation tools exist. These tools have been created with the objective of accurately reflecting the operation of the physical system, at high spatial and temporal resolution; however, the behavior of cyber infrastructure is not reflected. Despite the existence of simulation tools for cyber aspects such as computing and communication, differences in temporal resolution and data representation and lack of well-defined interfaces exert significant challenges to linking these simulation tools and providing an integrated CPS simulator.

For the electric delivery system, there are a number of commercial and non-commercial computer simulation tools available. The PowerWorld Simulator [88] is a popular commercial tool for analysis of high voltage power systems. It supports common protection and control devices, provides an interactive environment and intuitive GUI, and is able to solve power flow equations for very large systems; however, PowerWorld does not provide the transparency needed for analysis of the sequence of failures. Several other commercial software packages, such as DIgSILENT [89], have the same shortcoming. Among the non-commercial packages, MATPOWER [90] and PSAT [18] are two MATLAB-based toolboxes commonly used for academic research. MATPOWER can solve load flow and optimal power flow problems in a command line interface. PSAT has a graphical interface and supports basic monitoring and protection devices and power regulators in addition to the capabilities of MATPOWER. In this research, we used PSAT for simulation of the power grid systems. For the purpose of our simulations, we enhanced PSAT in order to achieve the high resolution required for analysis of smart grids. These enhancements include defining data structures for each category of cyber components, incorporating wide-area measurement capabilities by PMU devices, providing a platform for implementing a decision support algorithm, and integrating the power systems with communication technologies used in smart grid applications. This modified version of PSAT is interfaced with a MATLAB wrapper that acts as an adapter between libraries and orchestrates subroutine calls. Figure 4.5 visualizes the operation of our smart grid simulator.

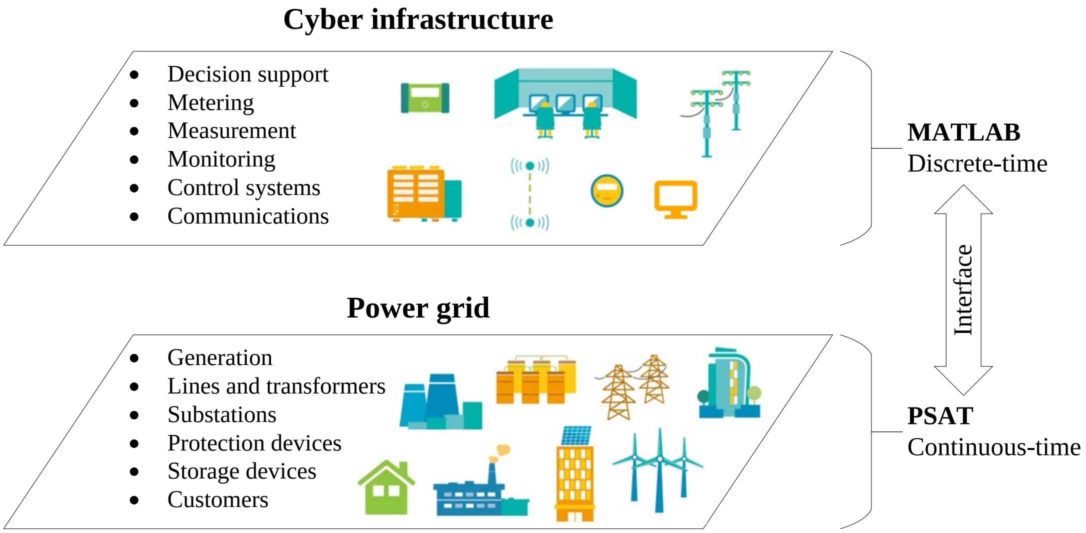


Figure 4.5. Cyber-physical smart grid simulator.

Inputs to the simulator include i) a data file that comprises power grid topology and cyber infrastructure specifications and ii) a perturbation file that lists all contingencies intended to be analyzed. On the other hand, the simulator returns system state, including electrical parameters (e.g., voltages, phases, and power flows) as well as operation status of cyber components. This output can be used to find the service indices and determine wether the system is considered to be in a failed or operational state.

4.2. ANALYSIS OF INTERDEPENDENCIES

Smart grids are composed of several different subsystems working together to achieve a common goal. The close interactions between different entities, captured by the concept of interdependence, are due to the fundamental attributes of CPSs as well as the nature of the electric delivery system. In this section, we will study the interdependence in smart grids using our proposed identification and quantification techniques.

In addition to the intrinsic dependencies between components, we assumed that the operation of PMUs depends on the underlying power grid, i.e., a PMU device is disabled as soon as a voltage violation occurs at the bus on which it is installed. Voltage violation is defined to be outside of 0.9 to 1.1 per-unit[1] range, according to the EN-50160 standard [91].

4.2.1. Selection of Failure Cases. Selection of failure cases and determining the minimum number of failure cases needed for obtaining all of the dependencies are of great importance in the presented method. In general, the larger the number of failure cases is, the more accurate the model becomes; however, exhaustive examination of failure cases is infeasible for large systems. The study presented in [75] provides a method for selection of failure cases in analysis of power grids while a given accuracy is maintained. In this work, we analyzed the following scenarios:

- One or two simultaneous transmission line outages,
- at most one failed SSSC device,
- at most one failed PMU, and
- failure of the decision support algorithm.

It is worth mentioning that upon availability of simulation environments capable of modeling the communication infrastructure with high resolution, considering the effects of respective impairments will improve the quality of the model. Unless a sophisticated

[1] The per-unit representation denotes normalization by a base value.

model for channel impairments is utilized, considering communication failures simply adds redundant failure cases and complicates representation of the results without actually capturing the behavior of the communications. For interdependency analysis of IEEE-14 and IEEE-57 smart grids, we have not considered the communication failures.

Table 4.2 lists the number of simulations carried out for each test case. The total number of simulations for each system is product of the number of failure cases shown for each category of component. Note that in all of the failure cases at least one transmission line is tripped since it is observed that the system does not degrade otherwise, even in the presence of cyber faults; however, having one or more transmission lines tripped, impairment of cyber components can exacerbate the situation and lead to further degradation.

Table 4.2. Number of simulated failure cases.

	IEEE-14	IEEE-57
transmission lines	$\sum_{k=1}^{2} \binom{20}{k} = 210$	$\sum_{k=1}^{2} \binom{80}{k} = 3,240$
SSSC devices	$\sum_{k=0}^{1} \binom{3}{k} = 4$	$\sum_{k=0}^{1} \binom{7}{k} = 8$
PMU devices	$\sum_{k=0}^{1} \binom{3}{k} = 4$	$\sum_{k=0}^{1} \binom{12}{k} = 13$
decision support	$\sum_{k=0}^{1} \binom{1}{k} = 2$	$\sum_{k=0}^{1} \binom{1}{k} = 2$
total number of simulated cases	6,720	673,920

In this study, failure cases were selected based on the failure rate of components. Transmission lines were selected because they have a relatively high rate of failure and are a major source of power outages [81]. Additionally, we selected SSSC and PMU devices and the decision support as representative cyber components and because their failure can impact the state of the physical components. Including other components of the system can enhance the model and improve the accuracy of results.

4.2.2. Interdependencies of IEEE Bus Systems. In this section, we present interdependencies of IEEE-14 and IEEE-57 smart grids identified using both correlation and causation analyses. Correlation analysis requires state variable for each component to be defined. Equation (4.1) shows definition of state variables for each category of components in smart grids.

$$
\begin{aligned}
X_{L_{i-j}}(t) &= |\text{Active power flow of } L_{i-j} \text{ in p.u. at time } t| \\
X_{F_{i-j}}(t) &= \begin{cases} 1, & F_{i-j} \text{ is operational at time } t; \\ 0, & \text{otherwise.} \end{cases} \\
X_{P_i}(t) &= \begin{cases} 1, & P_i \text{ is operational at time } t; \\ 0, & \text{otherwise.} \end{cases} \\
X_{DS}(t) &= \frac{\text{Number of observable buses and lines}}{\text{Total number of buses and lines}}
\end{aligned}
\tag{4.1}
$$

In Equation (4.1), $X_{L_{i-j}}(t) \in \mathbb{R}^+$, $X_{F_{i-j}}(t) \in \{0, 1\}$, $X_{P_i}(t) \in \{0, 1\}$, and $X_{DS}(t) \in [0, 1]$. Note that $X_{DS}(t)$ is the portion of the power system that is observable to the decision support and is used as a measure since it well captures its operation and data dependency on PMU devices.

Figure 4.6 helps to see how these state variables can capture the operation of a smart grid during a failure sequence. In Figure 4.6, state variables of components of IEEE-14 smart grid during a selected failure sequence are plotted on a single horizontal axis. Note that the state variables are shown only for components that experience degradation. Each row presents state variable of a component and the rows are ordered according to the propagation of the faults, i.e., the two topmost rows correspond to the components whose failure initiates the failure case (L_{1-5} and P_2); third and fourth rows represent state variable of the components that fail consequently (L_{1-2} and F_{1-5}) and so forth. The bottommost row shows the state variable of component F_{2-3}, which is known to be the last component to be affected in this failure sequence.

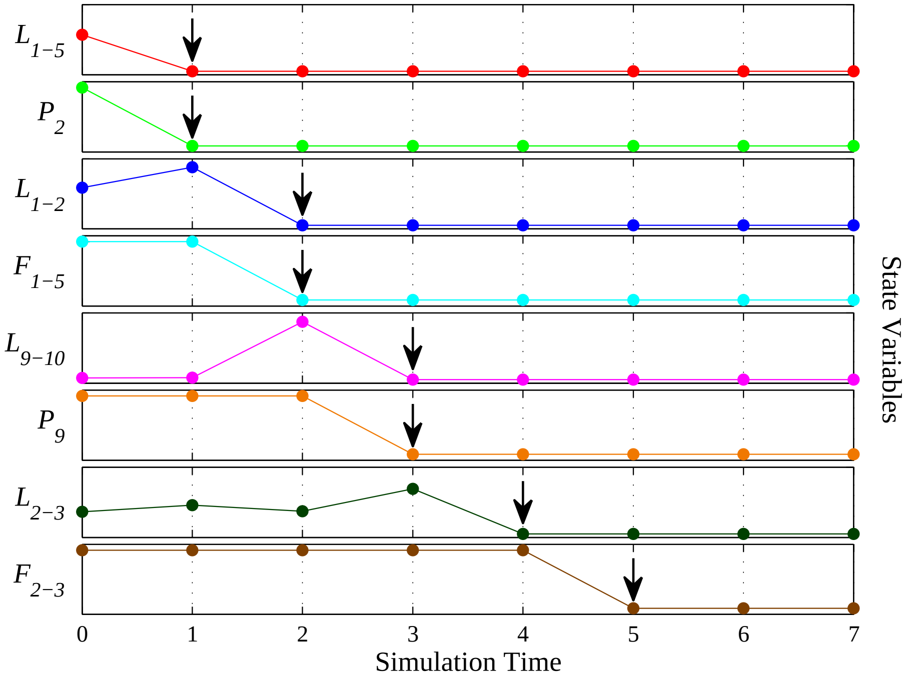


Figure 4.6. State variables of selected components of IEEE-14 during a failure sequence. Arrows indicate the points at which components are considered failed.

According to Figure 4.6, correlation among the state variables is expected to be maximal for those degrade with one time step difference. Table 4.3 shows values of *PCC* and *RDC* for pairs of components that degrade with one time step difference as shown in Figure 4.6. In calculation of *RDC* values, sample size and number of random features are set to 0.1 and 1, respectively.

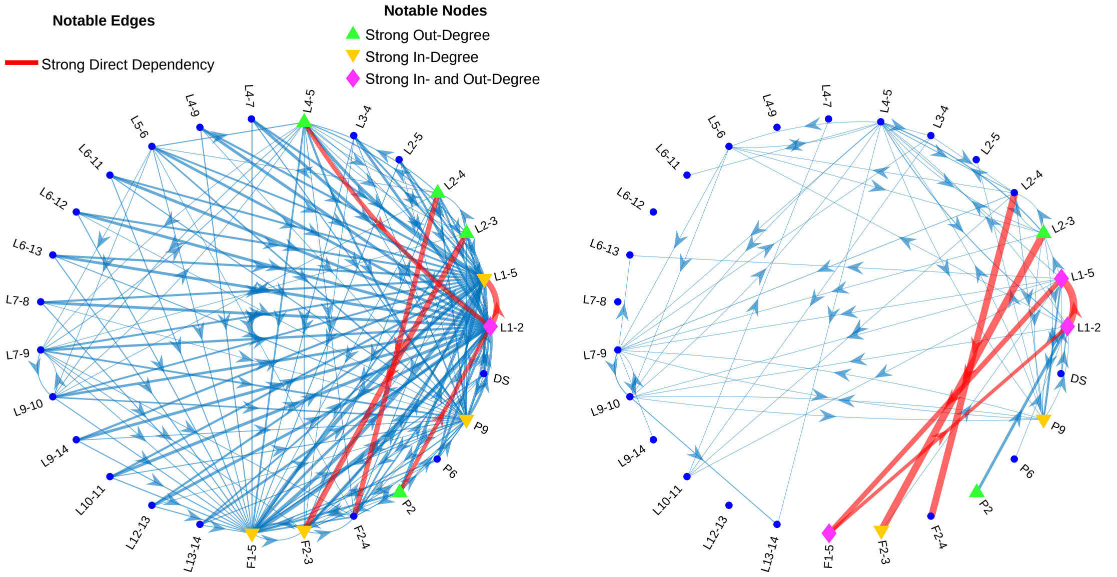
Both correlation coefficients exhibit relatively large values for the component pairs that are expected to have dependence; however, *RDC* values of dependent components better stand out according to Table 4.3. This is mainly due to the fact that *RDC* captures nonlinear as well as linear correlations, unlike *PCC*, which only captures linear relationships and results in underestimating dependency between component pairs that are non-linearly correlated. Hereinafter, we utilize and present *RDC* values only, due to its superiority in capturing the interdependence.

Table 4.3. Correlation coefficients between state variables of component pairs for those experience degradation with one time step difference in the failure case shown in Figure 4.6.

State variables		*PCC*	*RDC*
$X_{L_{1-5}}$	$X_{L_{1-2}}$	1.00	1.00
$X_{L_{1-5}}$	$X_{F_{1-5}}$	1.00	1.00
X_{P_2}	$X_{L_{1-2}}$	1.00	1.00
X_{P_2}	$X_{F_{1-5}}$	1.00	1.00
$X_{L_{1-2}}$	$X_{L_{9-10}}$	0.85	1.00
$X_{L_{1-2}}$	X_{P_9}	0.98	1.00
$X_{F_{1-5}}$	$X_{L_{9-10}}$	0.71	1.00
$X_{F_{1-5}}$	X_{P_9}	1.00	1.00
$X_{L_{9-10}}$	$X_{L_{2-3}}$	0.79	0.99
X_{P_9}	$X_{L_{2-3}}$	0.95	0.99
$X_{L_{2-3}}$	$X_{F_{2-3}}$	0.94	0.99
Maximum among all other pairs		0.68	0.83

For each failure case, *RDC* values are calculated for all pairs of state variables. For estimating d_{ij}, mean value of $RDC_{X_iX_j}$ is calculated over all failure cases. Likewise, the process explained in Section 3.1.2.2 is applied on the results of simulations to find the d_{ij} values using causation analysis. For each of the methods, we found d_{ij} values and constructed the **D** matrix.

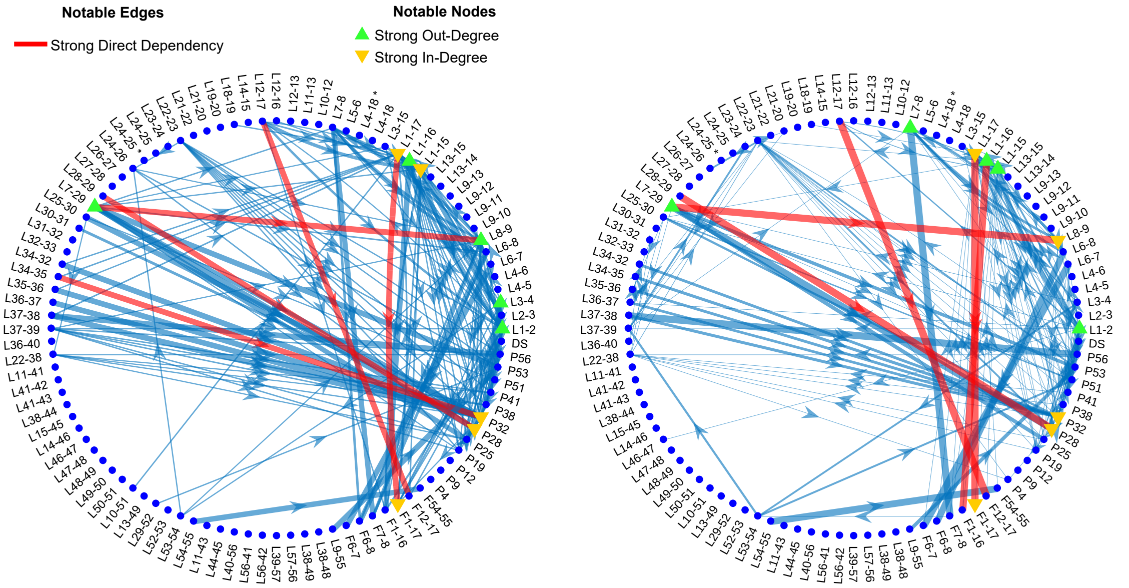
The **D** matrices are visually represented as weighted directed graphs, in Figure 4.7 and Figure 4.8 respectively, for IEEE-14 and IEEE-57 smart grid systems. Each edge represents a direct dependency and its width is proportional to the extent of the dependence (d_{ij}). Note that edges with weights less than 0.05 are not shown in the figures for ease of illustration. The top five links with largest weights are shown in red. It can be seen that dependency links with large weights are captured by both correlation and causation methods. It is also seen that the correlation-based method identifies a larger number of direct dependencies, as causation is typically a relationship that exists only in a small fraction of correlated events. In Appendix A, numerical values of notable direct dependency links are provided.



(a) Correlation (RDC)

(b) Causation

Figure 4.7. Graph representation of **D** matrix for IEEE-14 smart grid identified using correlation (a) and causation (b) analyses. Notable dependency links are shown in red.



(a) Correlation (RDC)

(b) Causation

Figure 4.8. Graph representation of **D** matrix for IEEE-57 smart grid identified using correlation (a) and causation (b) analyses. Notable dependency links are shown in red.

By applying interdependency quantification methods presented in Section 3.1.3, we computed **T** matrix based on direct dependencies identified using both correlation and causation analyses. Among the values of t_{ij} there exist links that are of great importance. As an example, in the IEEE-14 smart grid case, the total influence from the decision support to L_{1-5} is among the largest values while the corresponding direct link has a small weight of 0.05. The pair of P_{32} and F_{1-17} in the IEEE-57 smart grid is another example with a similar situation. This characteristic is justified by existence of several multi-step strong dependency links that connect pairs of components together and can give rise to further breakdown of components that are not in the geographical, logical, physical, or cyber reach of the initially impaired component [59]. This nonlocal property of the fault propagation has been observed in a number of real-world blackouts [2, 3]. Appendix A provides numerical values of notable links with large multi-step dependencies.

In Figure 4.7 and Figure 4.8, the top five components (nodes) with highest in-degree (ν) and out-degree (τ) values are specified by distinguishable markers. This analysis identifies the cyber and physical elements that have the highest priority for further inspection and fortification if dependability is to be improved. Among the identified components are the bridge lines, which are responsible for transmitting the majority of the power from the generating buses to the load buses, and PMU and SSSC devices that are installed on critical locations of the power grid. Identifying these components using analytical methods can be very difficult or even impossible for large systems. A more inclusive list of notable components with large in-degree and out-degree values is shown in Appendix A.

In order to lessen the risk of dependability loss due to existence of strong in-degree values, we can increase capacity of respective components to prevent overload, replace them with more robust components, and utilize redundant components [20]. For mitigating the risk exerted by strong out-degree values, a diverse selection of components should be utilized [92]. Diversity involves the use of spatially, temporally, and functionally different alternatives and guarantees that the components will not be affected by the same disruptions.

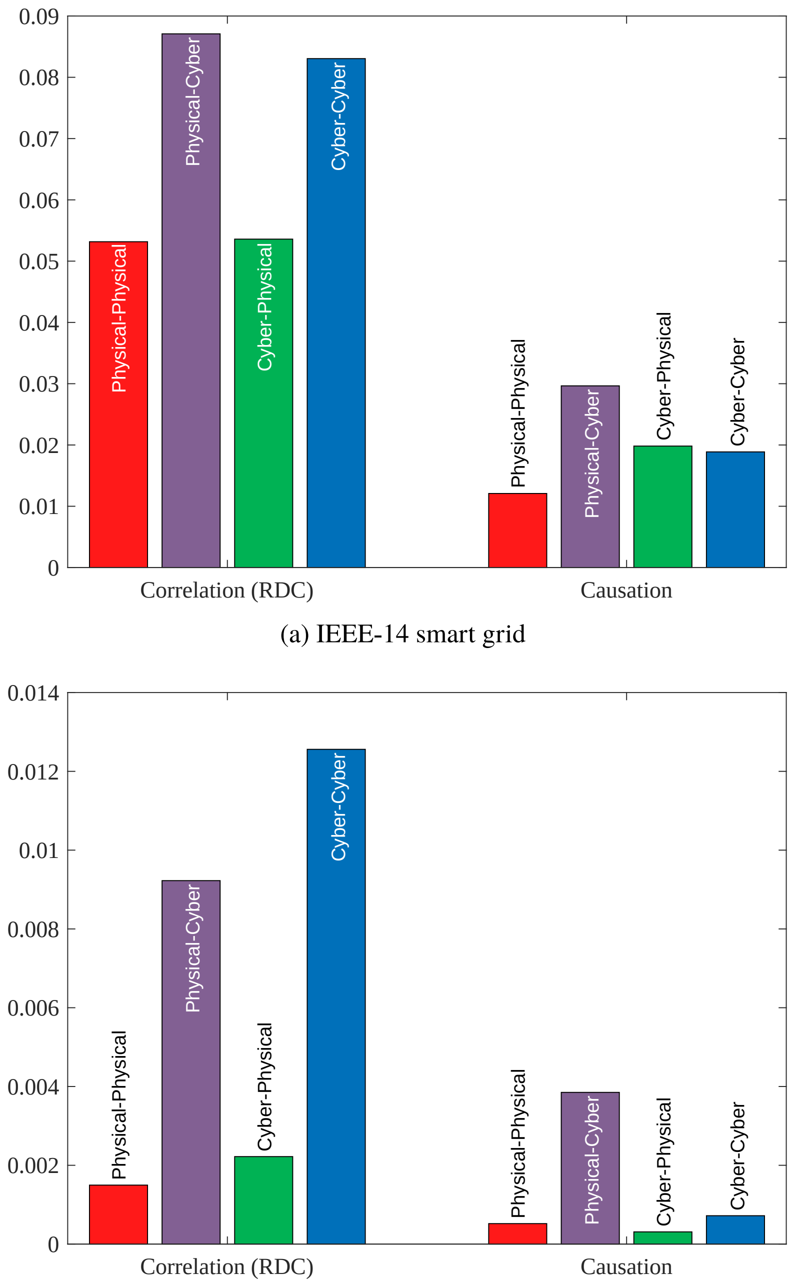
In order to compare the extent of dependency among the cyber and physical subsystems, $\gamma_{s_1-s_2}$ is calculated and shown in Figure 4.9. It can be seen that the most significant dependencies are from the physical subsystem to the cyber subsystem (physical-cyber) and within the cyber subsystem (cyber-cyber). Dependency of cyber components on the physical components are typically in binary form (e.g., if power is not available the sensor shuts off) which justifies the large value of corresponding $\gamma_{s_1-s_2}$ value. The large value of cyber-cyber dependency is mainly due to the nature of the common topologies of the cyber infrastructure, in which each component is connected to multiple other components for transmission and reception of information.

Comparing the respective $\gamma_{s_1-s_2}$ values obtained from the correlation and causation analyses, we can infer that the causation relationships among the components are less frequent and of smaller extent, which can also be observed in the graphs provided in Figure 4.7 and Figure 4.8; however, the orders of $\gamma_{s_1-s_2}$ values within each group are similar to each other.

4.3. RELIABILITY MODELING

The most commonly studied dependability attribute for the smart grids is reliability. This is mainly because of the fact that uninterrupted delivery of electricity has the highest priority among objectives of the smart grids and this capability is best captured by reliability. In this section, we apply our proposed reliability modeling approach to the smart grids. In addition, we measure the detrimental effect of introducing additional interdependency on the reliability of the IEEE-14 smart grid test case.

4.3.1. Definition of System Failure. Reliability is a dependability attribute that takes a binary view of the system state, and hence, it is very important to have a clear understanding of what is considered a "system-level failure." We define a system failure as comprising at least one of the following two cases:



(a) IEEE-14 smart grid

(b) IEEE-57 smart grid

Figure 4.9. Comparison of dependency among subsystems ($\gamma_{s_1-s_2}$) using correlation and causation analyses.

- Voltage violation: Having at least one load that is supplied by a bus whose voltage does not fall in 0.9 to 1.1 per-unit range (EN 50160 standard [91])

- Excessive outages: Concurrent failure of more than three transmission lines

Excessive outages is not a general criterion for system-level failure of smart grids, but according to extensive simulations on IEEE-14 and IEEE-57, leads to a cascading failure or a voltage violation eventually. It particularly helps in eliminating the need for exhaustive simulation of the entire system states.

4.3.2. Simulated Cases. A system composed of n components with binary states has a total of 2^n states; however, in most cases it is neither feasible, nor required to examine all these states. For IEEE-14 and IEEE-57 smart grids, it is verified that states with more than three concurrent transmission line outages will certainly lead to a failed state (also reflected in our definition of system failure) and do not need to be examined. Furthermore, occurrence of some states are practically improbable. Such states can be eliminated without a significant impact on the accuracy of the reliability model too.

Consequently, in this work we have analyzed the following scenarios:

- Up to three simultaneous transmission line outages,
- at most one failed SSSC device,
- at most one failed PMU,
- at most one failed communication channel, and
- failure of the decision support algorithm.

For the sake of brevity, results are presented with the assumption that components have binary states (i.e., operational or failed). Furthermore, failure of a component is considered to be representative of failure of all backups (e.g., backup batteries and generators and redundant communication mediums). Note that this assumption is made only for the examples of IEEE-14 and IEEE-57 smart grids and does not restrict the approach in general. With these assumptions, the proposed method is not computationally intensive. Recorded computational time for simulation of the IEEE 57-bus system is 22 minutes on an Intel Xeon E5-2623 3.00 GHz machine, which is acceptable considering that model composition is typically a non-real-time task.

4.3.3. Reliability Models for IEEE Bus Systems. If no cyber infrastructure is deployed for the IEEE 14-bus system, only two specific single-line contingencies cause system failure: respective outage of the transmission lines L_{1-2} and L_{1-5}. Any other single-line contingency leaves the system in an operational state. Furthermore, among the 190 possible double-line contingencies, only 44 leave the grid in a "failed" state. Equation (4.2) represents system-level reliability for the IEEE-14 when no cyber infrastructure is utilized, i.e., the system shown in Figure 4.3 without the decision support, SSSC and PMU devices, and the interconnecting communication links.

$$R_{sys} = p_L^{20} + 18q_L.p_L{}^{19} + 146q_L^2.p_L{}^{18} \tag{4.2}$$

In Equation (4.2), p_L and q_L, respectively, denote the reliability and unreliability of transmission lines, i.e., $q_L = 1 - p_L$.

For reliability analysis of the IEEE-14 smart grid, we need to investigate the system-level effect of all component-level failures, whether the component is in the physical or cyber infrastructure. Cyber devices can fail in various modes, ranging in effect from *fail-fault* - where failure of the cyber device severely impacts the operation of any other components dependent upon it, to *fail-bypass* - where the cyber device fails to zero, i.e., failure of the device is equivalent to removing it from the system. In fail-bypass mode, a failed PMU does not send any data to the decision support, a failed SSSC device does not adjust the power flow, and a failed decision support algorithm does not tune the SSSC devices. For brevity, we present only the results for a *fail-bypass* mode. Emphasis on fail-bypass mode is further justified by the relatively nascent stage of cyber-physical critical infrastructure, which motivates conservative deployment of cyber devices.

Figure 4.10 illustrates the overall system reliability of the IEEE-14 smart grid, as a function of transmission line reliability. The horizontal axes of the plots are in logit scale, chosen to emphasize common reliability values, expressed in "N nines" notation. Each subfigure captures the effect of improvements to component-level reliability of a specific category of cyber devices, i.e., the decision support algorithm, SSSCs, PMUs, and communication links, respectively. p_F, p_P, p_{CM}, and p_{DS}, respectively, denote reliabilities of each SSSC device, each PMU device, each communication link, and the decision support. Mathematical representation of the reliability model for IEEE-14 smart grid is shown in Appendix C.

Figure 4.10 confirms that more reliable transmission lines result in higher system-level reliability. Note that the high (e.g., five nines) reliability expected of a critical system is not achieved unless the transmission lines are highly reliable [93]. For the case study of a smart IEEE-14, assuming that no cyber device is less than 90% reliable (a conservative estimate), grid reliability of five nines requires transmission lines with a reliability of at least six nines. However, fortification focused solely on improving the reliability of transmission lines exhibits diminishing returns.

Figure 4.10 also facilitates comparison of the improvement achieved in system-level reliability as a result of enhancing the reliability of each category of cyber components. This information can be used to guide the investments in fortifying the grid. For example, comparing Figure 4.10b and Figure 4.10d confirms the intuitive notion that investing in more reliable communication links is much more rewarding than improving reliability of the SSSC devices.

Figure 4.11 highlights the "break-even" point for reliability of a smart grid, i.e., the minimum reliability required of cyber components for the IEEE-14 smart grid to be more reliable than its purely physical counterpart. This threshold is compared for different values of transmission line reliability. For simplicity, all cyber components have been assumed equally reliable. Among the four cases illustrated in Figure 4.11, the lowest threshold

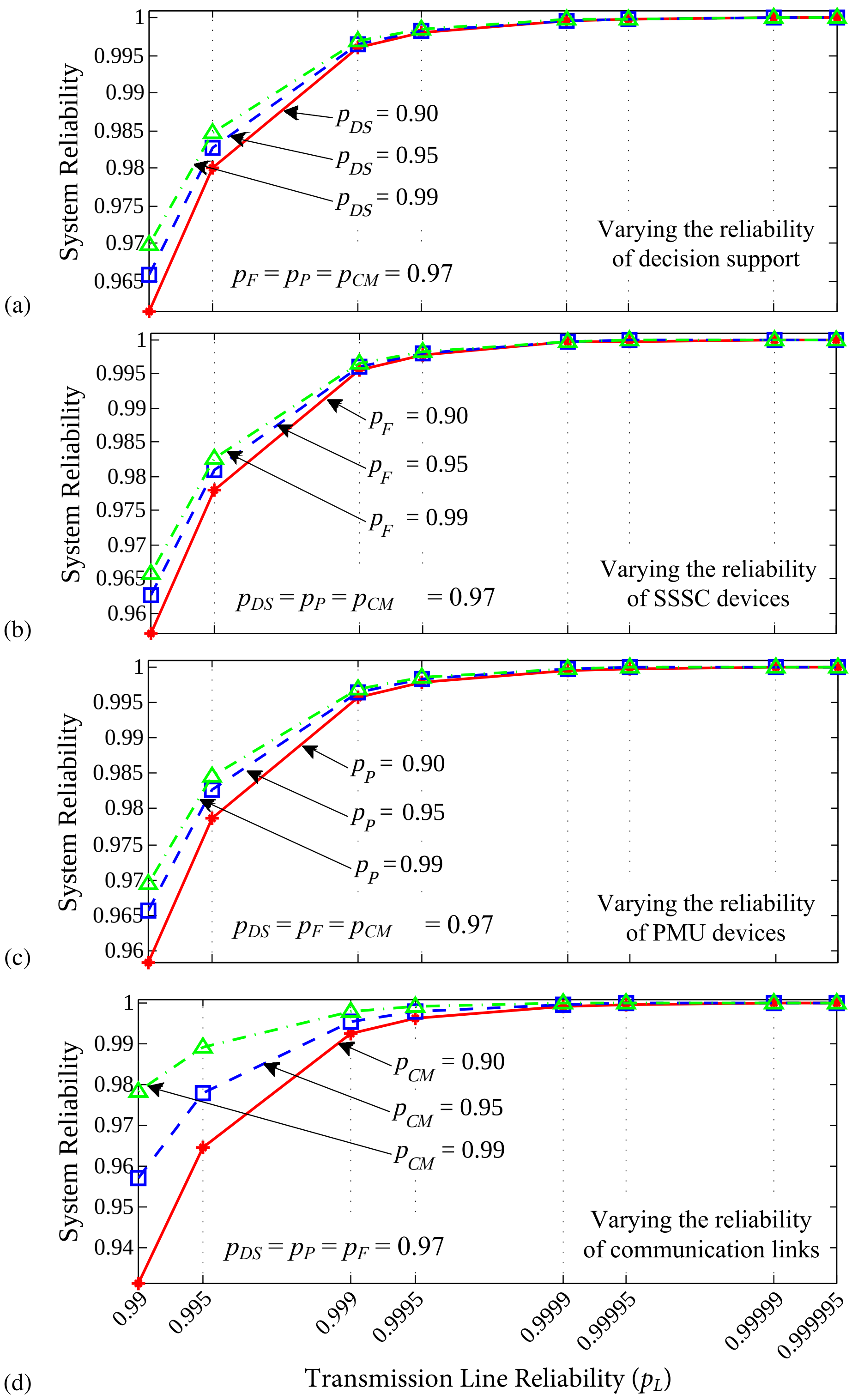


Figure 4.10. Effect of improving the reliability of cyber components on overall reliability of IEEE-14 smart grid. All subfigures share the horizontal axis of (d).

identified for the reliability of cyber components is 0.985 - relatively high for computing or communication equipment. Our reliability analysis for this case study assumed fail-bypass behavior for all cyber components. Other failure modes could amplify the effect of their unreliability and necessitate even higher thresholds for the reliability of cyber components.

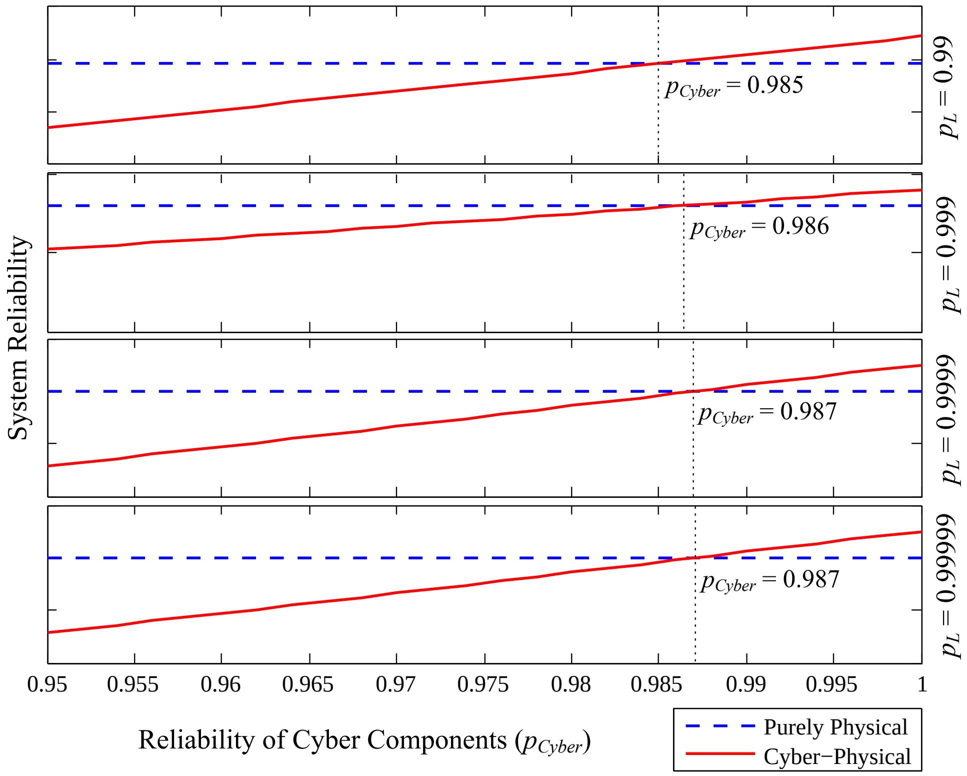


Figure 4.11. Comparison of the reliability of the smart and purely physical IEEE-14 grid.

For the larger case of IEEE-57 smart grid, overall system reliability shown in Figure 4.12 is more sensitive to the reliability of transmission lines. In general, the larger the size of the system, the more reliable its components need to be in order to maintain a specific level of overall system reliability. By comparing the relationship between the subfigures of Figure 4.12 with those of Figure 4.10, we can see that the cyber components in IEEE-14 and IEEE-57 smart grids have very similar effects on the reliability. This similarity is due to the fact that the cyber components in these two smart grids play similar roles, and confirms correctness of the inferences stated above about the effect of the cyber infrastructure on the reliability of smart grids.

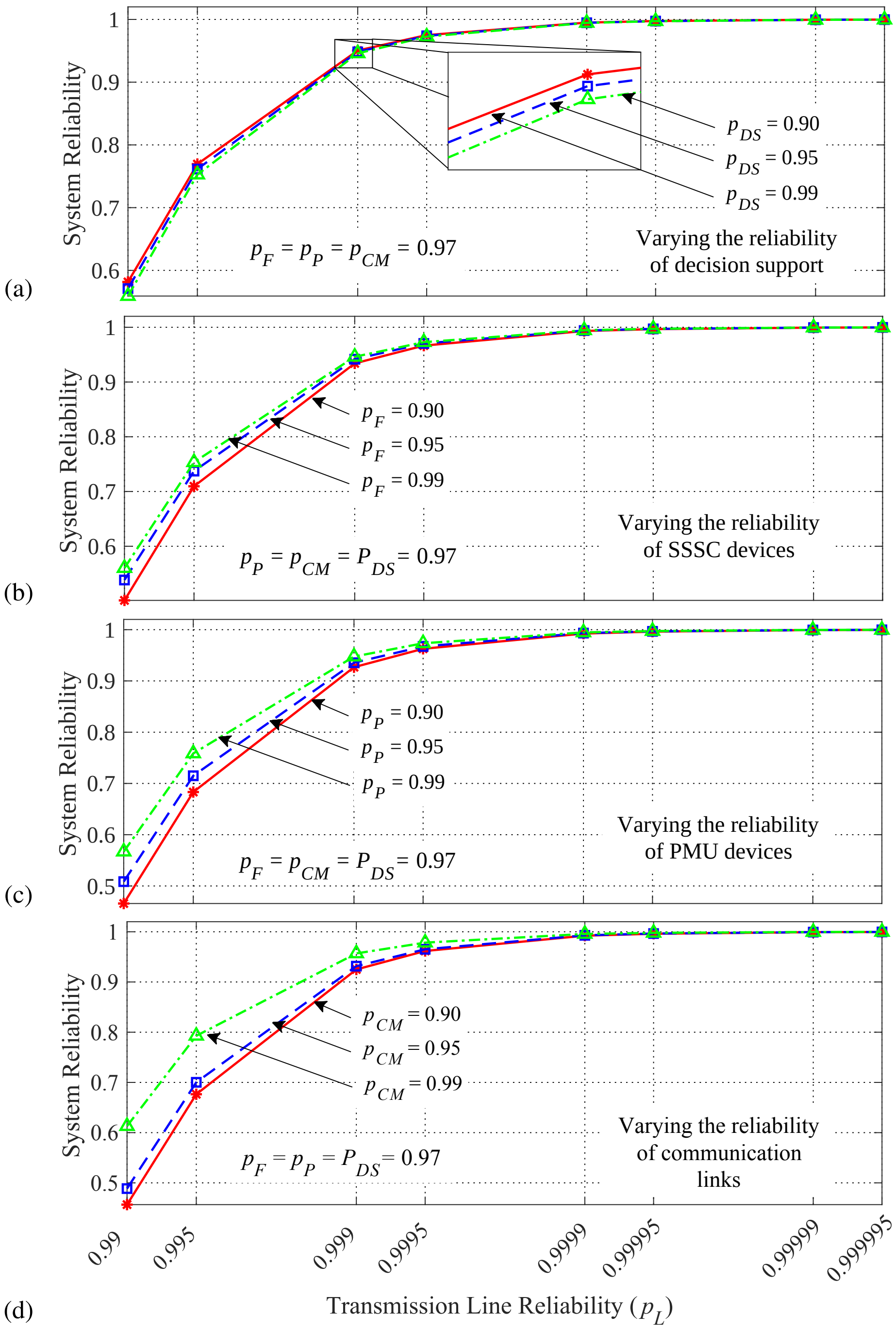


Figure 4.12. Effect of improving the reliability of cyber components on overall reliability of IEEE-57 smart grid. All subfigures share the horizontal axis of (d).

4.3.4. Effect of Additional Interdependence on Reliability. Cyber-physical interconnections among components of a CPS can expand the scope of information used for decision support and improve the system. It can also lead to additional fault propagation paths and degrade dependability. In this section, we investigate how added functional interdependency can degrade the reliability of the IEEE-14 smart grid, where cyber-induced interdependencies are of greatest interest. Our investigation is comprised of assuming additional dependencies - beyond those present in the original system - and evaluating their impact on reliability. Figure 4.13 compares the effects of adding dependencies of three respective types: physical-cyber, cyber-cyber, and cyber-physical. Transmission line reliability is held at 0.999. All cyber components, i.e., PMU and SSSC devices, communication links, and decision support, are assumed to have the same reliability value of 0.95. Each data point shown in Figure 4.13 represents the average reliability attained over 30 experiments, where each experiment consisted of modifying an arbitrary number of edges of the original dependency graph representing the IEEE-14 smart grid, while maintaining the value of $\gamma_{S_1-S_2}$. The weight of each modified edge, which represents the degree of influence of one component on the other, was arbitrarily chosen. To facilitate observation of trends, an exponential curve (depicted as a dotted line) was fitted to the data points for each type of dependence. In all three cases, system-level reliability degrades very quickly as the dependency index increases. Similar trends have been reported for degradation of robustness as connectivity increases [94].

Of the three types of dependency examined, cyber-physical dependencies, where failure of a cyber component brings down a physical component, were found to be the most crucial. Cyber components are typically more complex, and hence, cyber failures occur more frequently. System-level reliability is evaluated based on physical manifestations of failure, which directly reflect the operational state of physical components. Taken together, these two facts provide tangible justification of the rapid degradation observed in system-level reliability as cyber-physical dependencies increase.

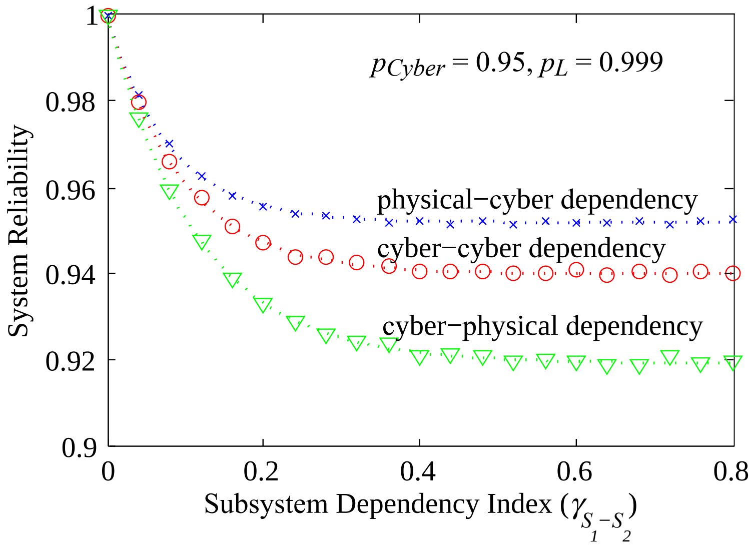


Figure 4.13. Comparison of the effect of physical-cyber, cyber-cyber, and cyber-physical dependency on the reliability of IEEE-14 smart grid.

4.4. SURVIVABILITY EVALUATION

It is very unlikely in a large-scale smart grid that at a given time period all of the components maintain a fault-free operation and the system remains in a perfect state. While reliability takes such optimistic view of the system operation, survivability characterizes degraded states as well. In this section, we demonstrate our proposed approach for survivability evaluation by applying it to smart grid test cases. On IEEE-57 smart grid, we also identify the components whose failure is the most detrimental to survivability. For verification of this importance analysis, we fortify the selected components and reevaluate the survivability. It is expected that the IEEE-57 smart grid with fortified components possess a higher survivability.

4.4.1. Selection of the Figure-of-Merit. The *essential service* expected of a smart grid is provision of stable power to customers. We define two corresponding FoMs: the customer service index and the average nominal voltage error. The *customer service index* (CSI) reflects the fraction of customers who have received this essential service, with a binary view – a customer is either served with adequate power or has not been served at all. In accordance with standards such as EN-50160 [91], our determination of whether a customer has been served is based on whether the voltage of the bus to which the customer is connected is within a predetermined range. For example, EN-50160 specifies a range of 0.9 to 1.1 per-unit. Equation (4.3) articulates the calculation of CSI.

$$\text{CSI} = \frac{\text{Number of customers served}}{\text{Total number of customers}} \tag{4.3}$$

The second FoM we propose for evaluating smart grid survivability is the *average nominal voltage error* (ANVE), which is calculated from the average voltage error over all load buses that experience undervoltage or overvoltage, as in Equation (4.4). An ANVE of 1 indicates that the grid is providing full service. In contrast with CSI, which solely reflects blackouts, ANVE considers brownouts as well.

$$\text{ANVE} = 1 - \frac{\sum_{i} |\text{Rated voltage at bus } i - \text{Actual voltage at bus } i|}{\text{Total number of customers}} \tag{4.4}$$

4.4.2. Selection of Failure Cases. As power grids are typically highly reliable and robust networks, most evaluations rely on $N-1$ or $N-2$ contingency analyses, i.e., a single failure or two concurrent failures. In this work, we analyzed the consequences of an outage of a transmission line or an SSSC device in the presence of a fault in the cyber network. The cyber faults injected to the smart grid are manifestations of data corruption: (i) incorrect data from PMUs, (ii) incorrect commands generated by the decision support algorithm, and (iii) undetected errors in the communications. Note that any one of these cyber faults alone can be tolerated by the system; however, if they are accompanied by an outage of a transmission line, further propagation of the failures is likely. Table 4.4 lists the simulated failure cases and the number of simulations carried out for each case.

4.4.3. Simulation Environment. Our smart grid simulator is used to determine power flows and voltages in the network during the failure cases. Figure 4.14 illustrates the procedure we have followed for simulating each failure case. In each outer loop, a data file

Table 4.4. Types and numbers of faults simulated.

	IEEE-14	IEEE-57
single transmission lines	20	80
SSSC devices	3	7
number of hardware faults simulated	23	87
PMU devices	3	12
communication links	6	19
control units	1	1
number of cyber faults simulated	10	32
total number of simulation runs	230	2,784

that contains the topology of the system under test is loaded and a failure case (with index k) is executed at time t_e by injecting corresponding faults and/or failures. In the inner loop, at each time step, PSAT performs power flow analysis and determines active power flow on each line and voltage at each bus. PMU devices then measure phasor data (including active power and voltage) of corresponding lines and buses and send it to the decision support algorithm where new settings for SSSC devices are calculated. Updated settings will regulate active and reactive power flow in the lines, where SSSC devices are installed. At this point, power flow analysis is run once more to find the updated active power flows and bus voltages. In every iteration of the inner loop after instant t_e, active power flow of the lines are compared to their capacity, and if any line is overloaded, it is considered failed and the topology is updated accordingly.

The simulation continues until no further failures are detected. For the sake of consistency among the two IEEE bus systems and ease of comparing the plots, all simulations are continued for 25 time steps (denoted as t_{final} in Figure 4.14), however, all failure sequences terminate before the 25th time step.

Note that since the time is discrete and is determined by the software simulation tool, the rate of degradation is bounded to a maximum value. Additionally, minor changes in the rate of degradation due to time-specific variations may not be captured in the simulation environment.

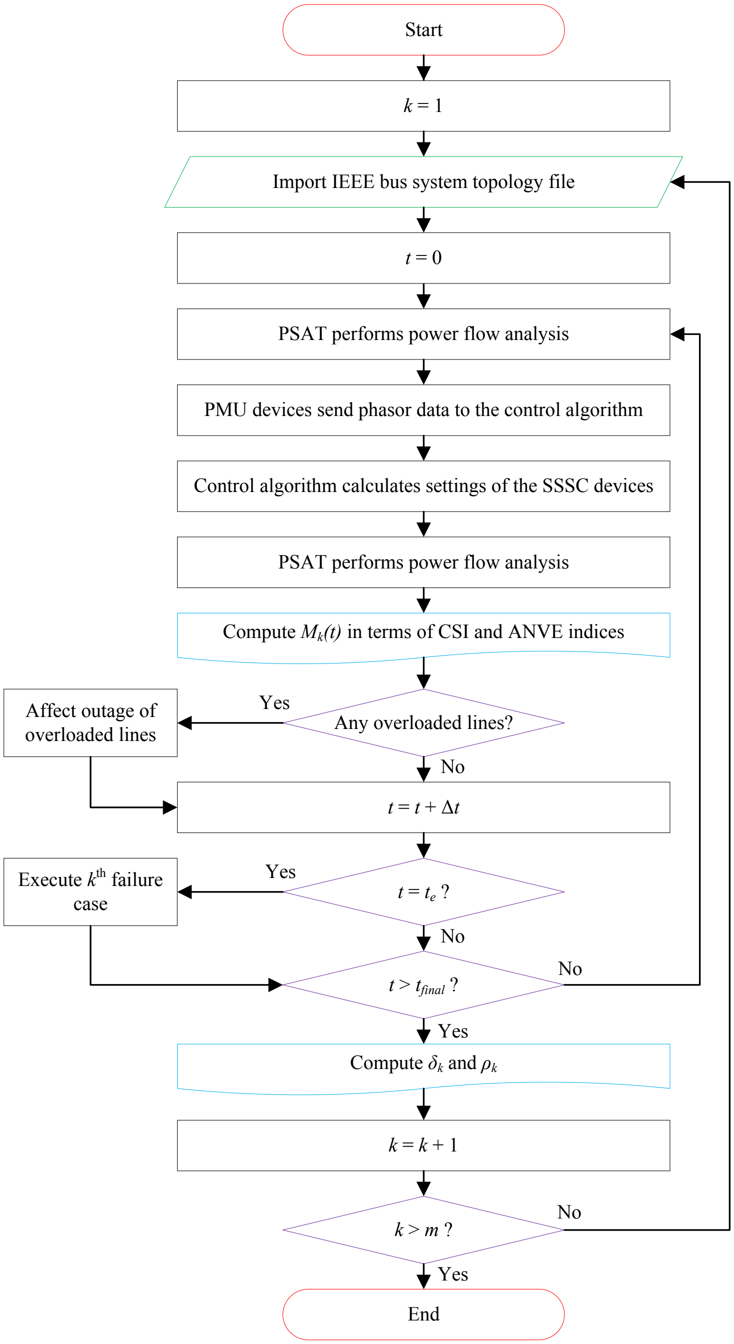
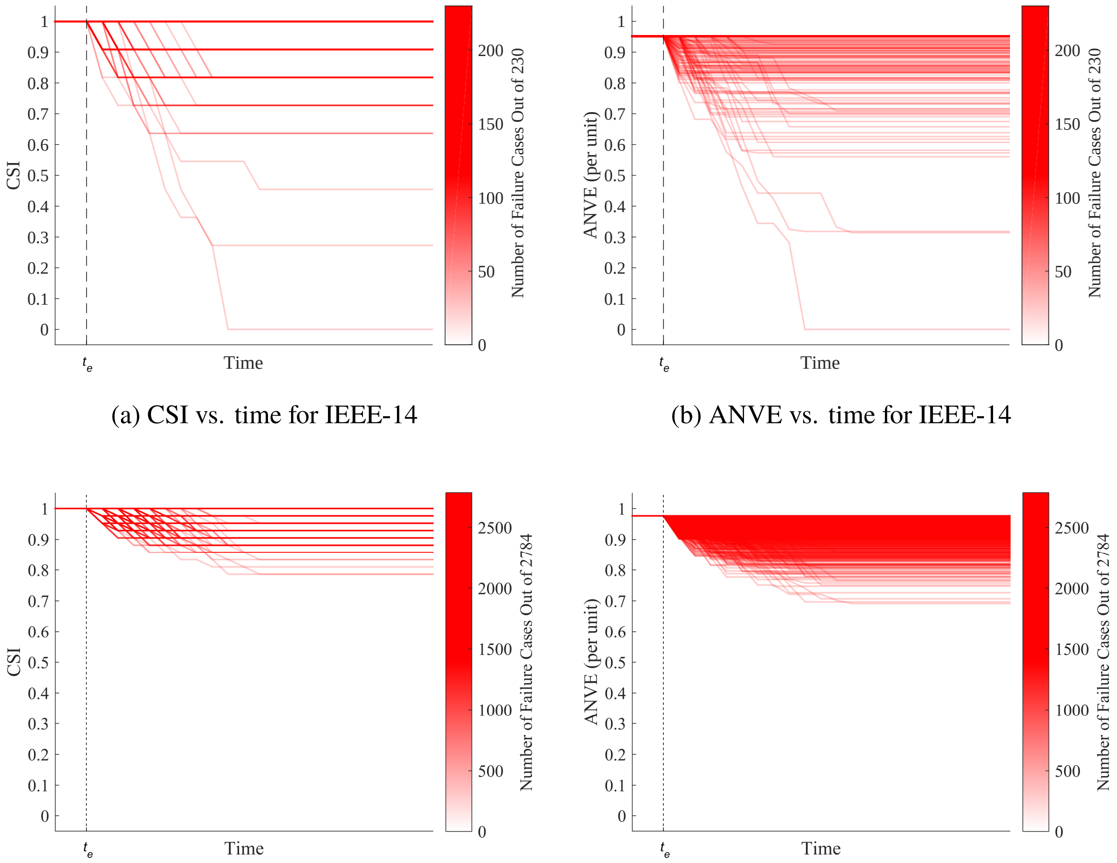

Start
$k = 1$
Import IEEE bus system topology file
$t = 0$
PSAT performs power flow analysis
PMU devices send phasor data to the control algorithm
Control algorithm calculates settings of the SSSC devices
PSAT performs power flow analysis
Compute $M_k(t)$ in terms of CSI and ANVE indices
Any overloaded lines?
Yes
Affect outage of overloaded lines
No
$t = t + \Delta t$
$t = t_e$?
Yes
Execute k^{th} failure case
No
$t > t_{final}$?
No
Yes
Compute δ_k and ρ_k
$k = k + 1$
$k > m$?
No
Yes
End

Figure 4.14. Survivability evaluation procedure.

4.4.4. Simulation Results. Figure 4.15 depicts the simulation results for each of the two test systems, using CSI and ANVE as the FoMs. In Figure 4.15, each sub-figure depicts the change in one FoM over time, after the injection of a failure. The intensity of a line indicates the number of failure cases which resulted in the behavior shown by that line. The desired outcome is a value of one, characterized by all customers being served and no voltage error for all customers, respectively, for CSI and ANVE. Note that since the CSI is discrete, it can hold only a finite set of values between 0 and 1.



(a) CSI vs. time for IEEE-14

(b) ANVE vs. time for IEEE-14

(c) CSI vs. time for IEEE-57

(d) ANVE vs. time for IEEE-57

Figure 4.15. CSI and ANVE vs. time for IEEE14 and IEEE-57 smart grids.

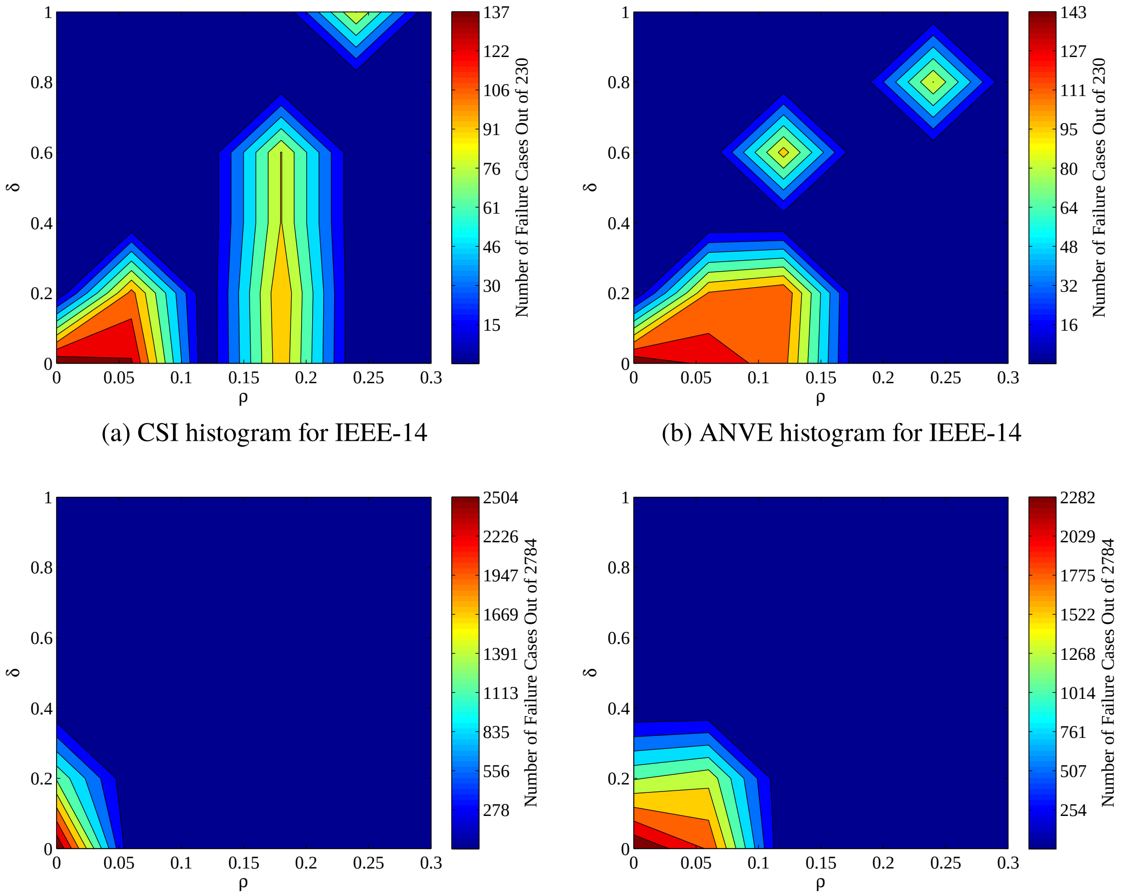
These results indicate that the majority of the simulated failure cases result in minimal degradation as the test systems are relatively robust. In the IEEE-14 smart grid results, shown in Figure 4.15a and 4.15b, a number of failures lead to total system failure with no customers served and maximum error for CSI and ANVE, respectively. This is indicated by the FoMs reaching zero. Additionally, the results show that a number of failure cases have two phases of system degradation separated by a brief period of stabilization. The IEEE-57 smart grid incorporates more redundancy and can tolerate a greater number of failures. This is seen in 4.15c and 4.15d where the FoMs never reach zero.

4.4.5. Evaluation of Survivable Behavior. The maximum rate and extent of degradation were extracted from the log of each failure case. Figure 4.16 depicts a two-dimensional histogram of CSI and ANVE for each smart grid system. These histograms are based on the maximum rate and extent of degradation calculated from the log of each failure case.

In an ideal system, every one of these histograms would be dense near the origin and sparse elsewhere, reflecting slow and minimal degradation in response to failure. This expectation is realized for the IEEE smart grids evaluated. However, there are clusters of failure cases with higher rates and extents of failure, which appear in the upper and/or right regions of the histogram. The presence of these clusters indicates that many of the failure cases simulated result in similar rates and extents of degradation. This is most likely caused by similar failure propagation paths through the power grid, i.e., different cascading failures involving the same vulnerable components.

4.4.6. Identifying Important Components. Our importance analysis technique is used to identify survivability bottlenecks and guide investments in fortifying these systems. Criticality and fragility can be determined for each component of a system, as described in Section 3.3.3.

Table 4.5 shows the rankings of the top ten lines of IEEE 57-bus system using fragility and criticality as criteria for hardening prioritization. It can be seen that some lines have similar ranking in both, e.g., lines L_{4-18}, L_{3-4}, and L_{4-6}. The also exist a few lines that have significantly higher priority using criticality as the metric, such as lines L_{8-9} and L_{6-7}, which is due to the difference in the weaknesses captured by fragility and criticality metrics. These lines fail in fewer failure cases (resulting in small fragility), but have very



(a) CSI histogram for IEEE-14

(b) ANVE histogram for IEEE-14

(c) CSI histogram for IEEE-57

(d) ANVE histogram for IEEE-57

Figure 4.16. Histograms showing extent of degradation vs. rate of degradation for CSI and ANVE. Color indicates the number of failure cases which resulted in the corresponding degradation point.

high impact on the system FoM when they do fail (as characterized by high criticality). Alternatively, a few lines have a significantly lower priority using criticality as a metric, such as lines L_{1-2} and L_{1-15}. These lines fail very frequently, but their failure is relatively insignificant in terms of system survivability.

4.4.7. Validation of Approach. In this section we validate our importance analysis technique through targeted hardening of the IEEE 57-bus smart grid. To harden this smart grid system the five lines with highest priority metrics were fortified by increasing their power flow capacity by 50%, which is expected to increase the survivability of the system as

Table 4.5. Transmission lines of IEEE 57-bus system with highest fragility and criticality. Only the top ten lines are shown.

Rank	Line	Fragility (×10)	Rank	Line	Criticality (×10^2)
1	L_{4-18}	0.280	1	L_{8-9}	0.148
2	L_{1-2}	0.273	2	L_{4-18}	0.142
3	L_{3-4}	0.251	3	L_{3-4}	0.134
4	L_{1-15}	0.237	4	L_{6-7}	0.128
5	L_{1-17}	0.237	5	L_{4-6}	0.114
6	L_{4-6}	0.223	6	L_{1-2}	0.108
7	L_{8-9}	0.223	7	L_{1-15}	0.106
8	L_{1-16}	0.223	8	L_{13-15}	0.072
9	L_{6-7}	0.216	9	L_{1-16}	0.066
10	L_{2-3}	0.194	10	L_{2-3}	0.064

it increases fault tolerance. The same hardening effect could have been achieved by adding redundant lines; however, this was not done in order to maintain the topology of the system for ease of comparison. Once the system was hardened the survivability analysis was rerun to compare the results with the original system.

First, fragility was used to select components for hardening. Lines L_{4-18}, L_{1-2}, L_{3-4}, L_{1-15}, and L_{1-17}, shown highlighted in yellow in Figure 4.17, were hardened. Next, criticality was used to select components for hardening. Lines L_{8-9}, L_{4-18}, L_{3-4}, L_{6-7}, and L_{4-6}, shown highlighted in blue in Figure 4.17, were selected to be hardened.

Comparison of the results of simulations as well as the survivability attributes for original and hardened versions of IEEE-57 (shown in Figure 4.18 and Figure 4.19) verifies effectiveness of the hardening technique. It is seen that for the hardened systems the extent of degradations has reduced and clusters of degradation points have moved towards the origin. Due to the choice of the hardening method, i.e., increasing the power flow capacity of lines, improvements in reducing the rate of degradation is not significant as seen in Figure 4.19. Other hardening methods, e.g., use of power storage and installation of protective relays, can be more effective in lowering the rate of degradation.

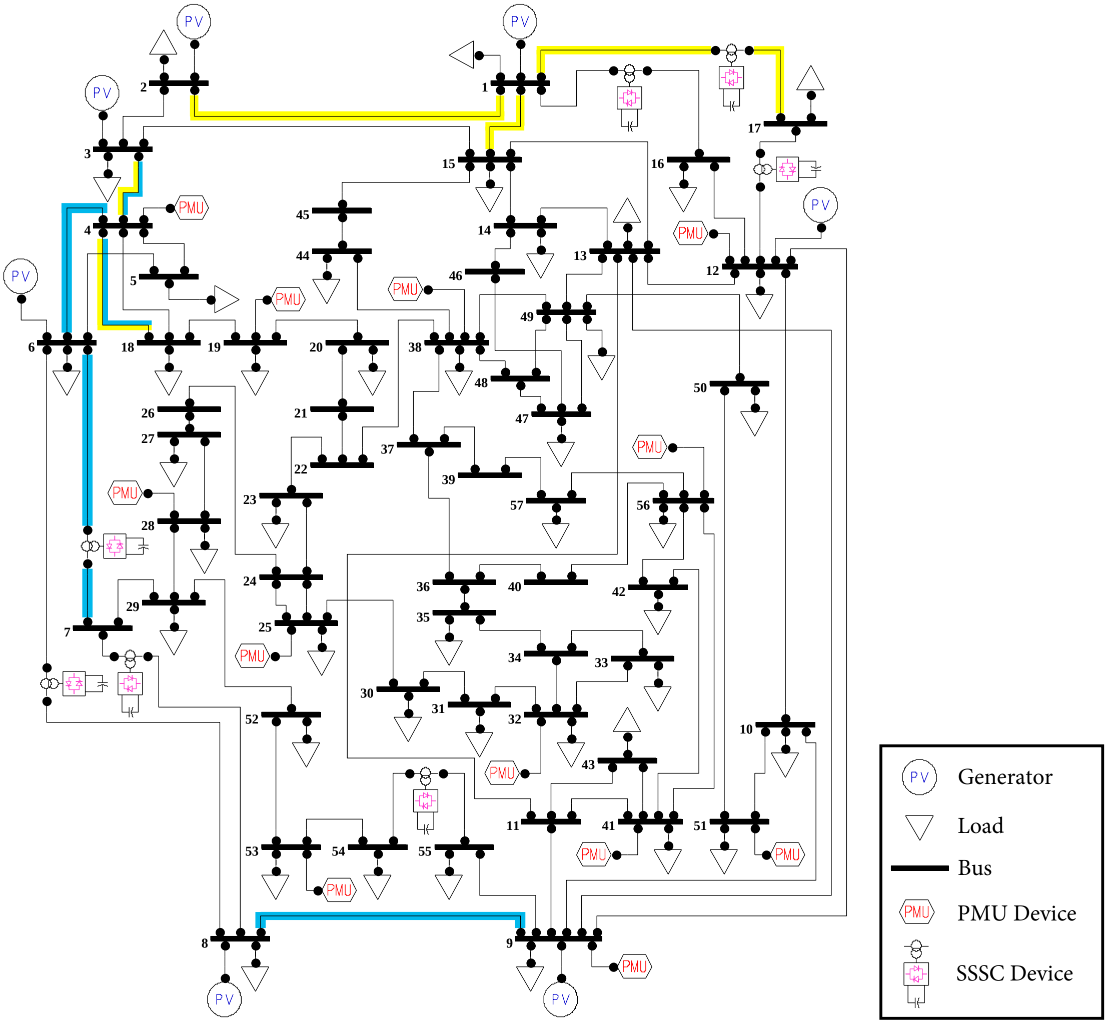
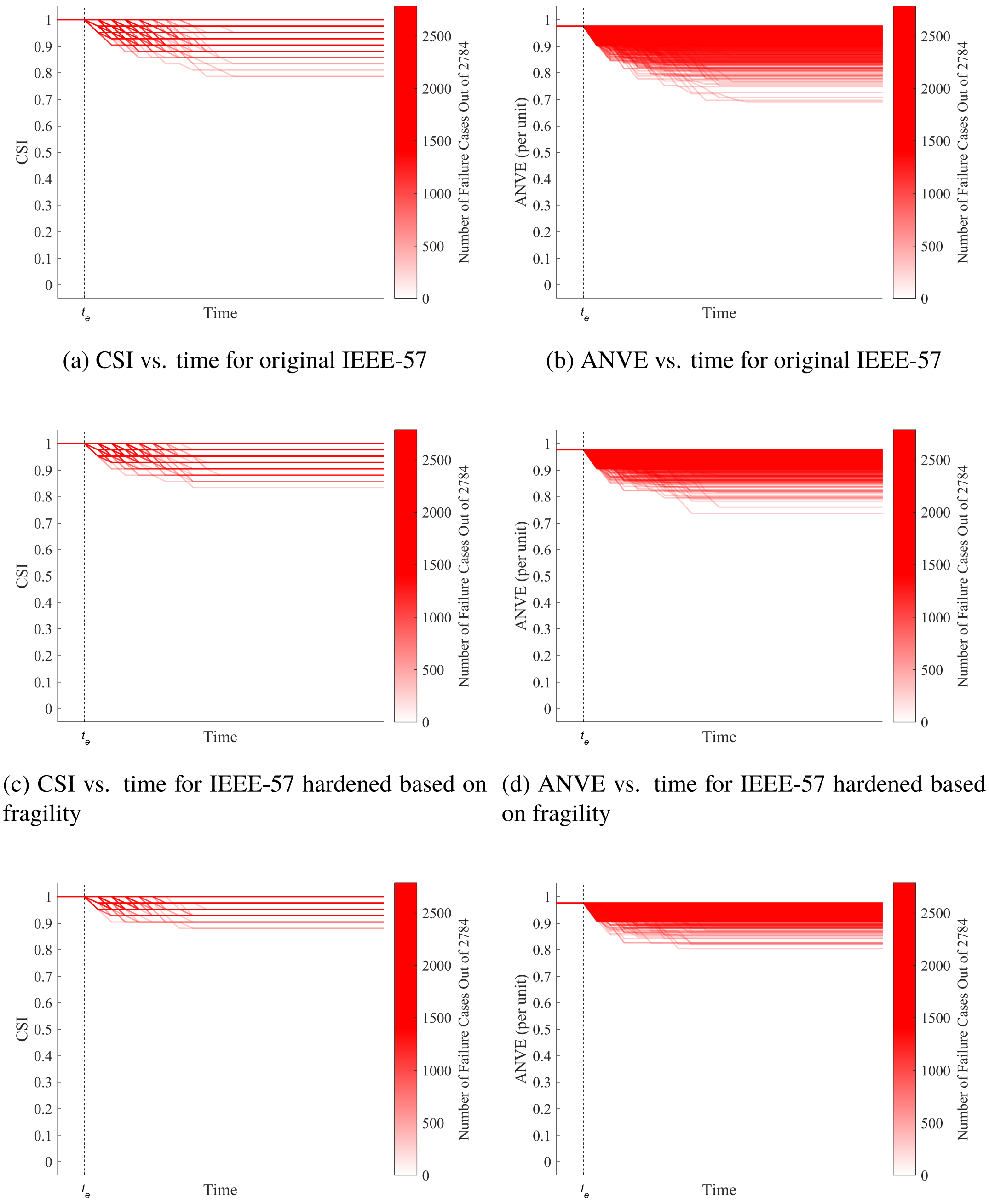


Figure 4.17. IEEE 57-bus smart grid test system. Lines highlighted in yellow have the highest fragility and those highlighted in blue have the highest criticality.

Comparing the survivability evaluation results of the two original and hardened IEEE 57-bus smart grids demonstrates an improvement in the survivable behavior of the system. Both importance analysis techniques resulted in an improvement in system survivability, using both FoMs, evident in Figure 4.18 and Figure 4.19; however, using criticality as the metric leads to a more effective improvement over the original system. This is due to the fact that the criticality metric better captures contribution of a component to survivability of a system, compared to the fragility.



(a) CSI vs. time for original IEEE-57

(b) ANVE vs. time for original IEEE-57

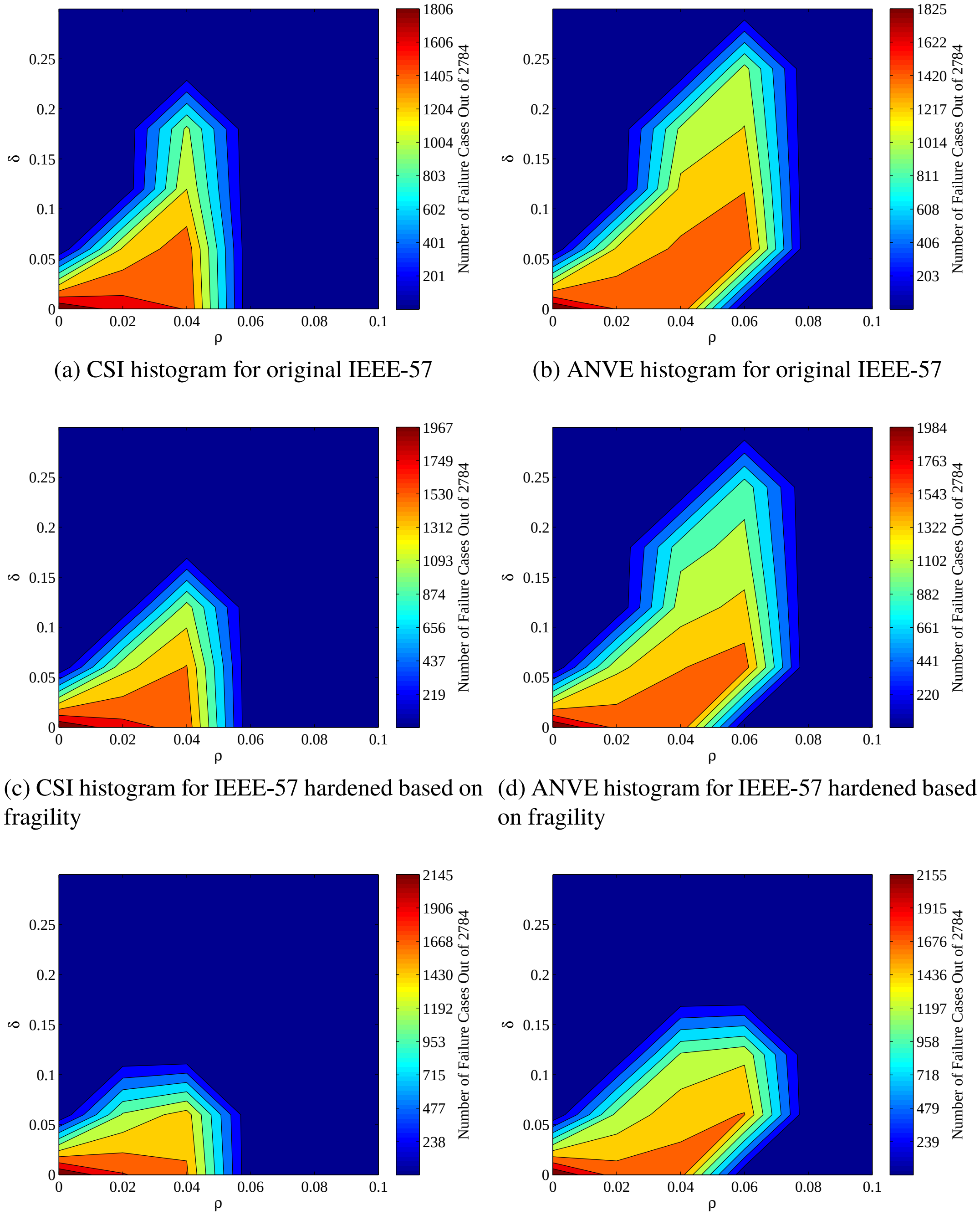
(c) CSI vs. time for IEEE-57 hardened based on fragility

(d) ANVE vs. time for IEEE-57 hardened based on fragility

(e) CSI vs. time for IEEE-57 hardened based on criticality

(f) ANVE vs. time for IEEE-57 hardened based on criticality

Figure 4.18. CSI and ANVE vs. time for original and hardened IEEE-57.



(a) CSI histogram for original IEEE-57

(b) ANVE histogram for original IEEE-57

(c) CSI histogram for IEEE-57 hardened based on fragility

(d) ANVE histogram for IEEE-57 hardened based on fragility

(e) CSI histogram for IEEE-57 hardened based on criticality

(f) ANVE histogram for IEEE-57 hardened based on criticality

Figure 4.19. Histograms showing extent of degradation vs. rate of degradation for CSI and ANVE of original and hardened IEEE-57.

4.5. PREDICTION OF FAILURES

In this section, we will present the results of training a failure prediction ANN using the data attained by simulating several failure cases on smart grid test cases and evaluate its performance on predicting imminent failures. Since the failure data for real-world large-scale CPSs is limited, we will investigate the possibility of training the ANN with a small subset of the available data sets and inspect its predictive capability. We randomly selected a subset of the available data and then divided them into training, validation, and test data sets using random selection. The training data set is used for adjusting the weights of the ANN shown in Section 3.4, using the backpropagation technique, while the validation data set is employed to minimize overfitting during the training cycles. Upon completion of the training, the test data set is used for measuring the predictive performance of the ANN.

We also simulated a number of randomly selected more complex failure cases for each smart grid system and evaluated the performance of the ANN on predicting failures on those cases. The difference between the "simple" and the "complex" test data is that the latter is composed of failure cases with three to five concurrent transmission line failures, while the former is selected from the failure cases explained in Table 4.2, where at most two transmission lines are tripped concurrently. Table 4.6 shows the number of entries of the failure data used for the ANN.

Table 4.6. Size of the failure data sets used for the ANN.

Data set	IEEE-14	IEEE-57
Total failure data available	17,968	1,181,871
Simple failure data used for the ANN	10,000	20,000
Training data (80%)	8,000	16,000
Validation data (10%)	1,000	2,000
Test data (10%)	1,000	2,000
Complex test data	1,000	1,000

4.5.1. Predictive Performance. Predictions on the failure cases from the complex data sets are expected to be harder for the ANN as they are not of the same type of the input data by which the ANN is trained. Performance measures of the ANN on simple and complex test data sets are shown in Table 4.7. It is worth mentioning that the failure prediction process, from inputting the system state to the ANN until receiving the output in the form of component indices that are about to fail, takes less than one millisecond on an Intel Xeon E5-2623 3.00 GHz machine.

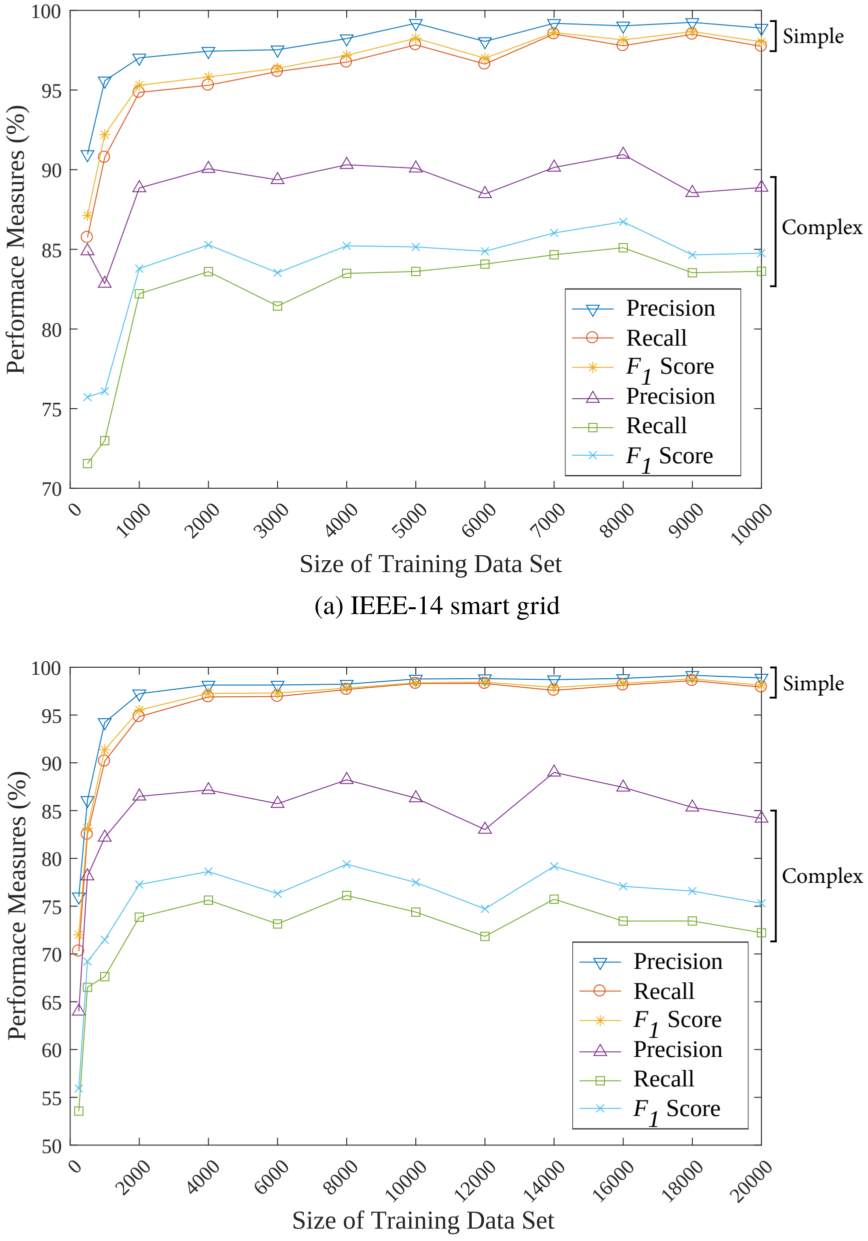
Table 4.7. Predictive performance measures of the ANN.

System	Test Data	Precision	Recall	F_1 score
IEEE-14	Simple	99.25%	98.21%	98.46%
	Complex	90.63%	83.65%	85.54%
IEEE-57	Simple	99.38%	98.66%	98.87%
	Complex	84.83%	71.55%	75.29%

As seen in Table 4.7, the ANN has an excellent performance on the simple data sets, both for IEEE-14 and IEEE-57 smart grids. Although the ANN does not perform as good on the complex data sets, its performance is yet acceptable. The fact that the ANN can predict imminent failures with a high accuracy is mainly in virtue of the interdependence among the components and existence of recurred failure sequences. Identification of such sequences of failure that frequently occur is also useful in fortification of the system [12].

4.5.2. How Much Training Data is Needed?. Thus far, we have seen that the ANN has a good performance in detecting the components that are about to fail, both with high accuracy and high speed. Another equally important feature of a prediction tool is that it maintains its performance when it is trained with a relatively small data set. In order to investigate whether the proposed approach has this feature and to find the minimum number of required entries in the training data set, we have performed the training process with subsets of the available data set and measured the predictive performance of the resulting

ANN on a fixed test data set. Figure 4.20 shows how the performance measures of the ANN in predicting failures of the IEEE-14 and IEEE-57 smart grids are affected by varying the size of the training data set.



(a) IEEE-14 smart grid

(b) IEEE-57 smart grid

Figure 4.20. The effect of the size of training data set on predictive performance of the ANN.

As expected, it is seen in Figure 4.20 that the performance of the ANN degrades as we decrease the size of the training data set. Note that due to the randomness in selection of the training data, the performance curves are not monotonically increasing. According to Figure 4.20, increasing the size of the training data set to more than 1000 for IEEE-14 smart grid and 2000 for IEEE-57 smart grid does not have a significant effect on the predictive performance of the ANN. Since the performance of the proposed ANN is not contingent on having large training data sets it can be trained using data available in the reports of the past power outages for developing a failure prediction tool for smart grids.

4.6. EXTENSION TO OTHER CRITICAL INFRASTRUCTURES

The essential assumptions made in this research are not domain-specific, making our approach applicable to various CPSs in spite of differences in commodities, e.g., water, power, vehicles, carried by each system. In fact, our modeling approaches seek to capture and quantify the behaviors that are described by the dependability attributes without relying on characteristics that are specific to the smart grid domain. As a result, our metrics and models can be generalized to other critical infrastructures, such as intelligent water distribution networks and intelligent transportation systems.

Prerequisites of applying the proposed techniques to other domains are: i) a simulator capable of capturing the operation of cyber and physical entities, ii) system descriptions and specifications such as physical and cyber topologies, and iii) definitions and assumptions regarding the acceptable operation of the system. Among these requirements, the most challenging problem is to build integrated cyber-physical simulators as further explained in Section 4.1.3.

4.6.1. Intelligent Water Distribution Network. A prominent example of critical infrastructure CPSs are intelligent water distribution networks (IWDNs), which are very similar in topology and structure to smart grids. The physical infrastructure of a smart grid is responsible for delivery of electricity from he generators, through transmission and distribution lines, to end-users. Similarly, an IWDN transfers water from reservoirs and tanks to customers through a network of pipes. Figure 4.21 depicts a hypothetical IWDN.

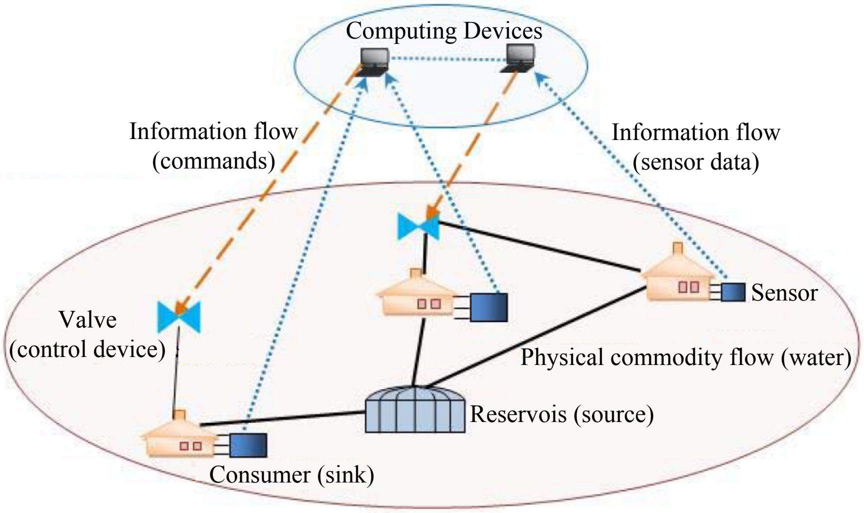


Figure 4.21. An IWDN.

The cyber infrastructure of the IWDN seeks to make effective use of water resources by increasing the number of pathways for water circulation. Decision support uses the information collected by several sensors dispersed in the physical infrastructure and actuates the hardware controllers to manage the allocation (quantity) and chemical composition (quality) of the water. Despite the undeniable performance gains facilitate by the cyber infrastructure, it is critical to verify that dependability aspects is not compromised. This task is facilitated by model-based approaches such as our proposed MIS model, which we derive for an IWDN in the remainder of this section.

4.6.1.1. Test case. The IWDN analyzed in our case study, depicted in Figure 4.22, consists of two water sources (a reservoir at node 1 and a tank at node 11), nine demand nodes (2-10), four valves (96, 97, 98, 99), and thirteen pipes (1-11, 98a, 98b). The reservoir is capable of providing an infinite supply of water, while the tank's supply is limited by the tank diameter and water level, which are predefined. The system has a single pump located at the reservoir which maintains the flow and head. The pipes and valves are the main components that control the flow and provide water to the consumers represented by the demand nodes. The topology of the physical water distribution network studied is based on the water network analyzed in [95, 96] to enable comparison of the reliability evaluation results. We added additional components, including a tank and multiple valves, to create a more robust system, but neither the basic structure, i.e., elevation of nodes and topology of the network, nor the supply and demand (in million gallons per day) specifications were altered. Table 4.8 presents the parameters associated to the physical infrastructure of this IWDN.

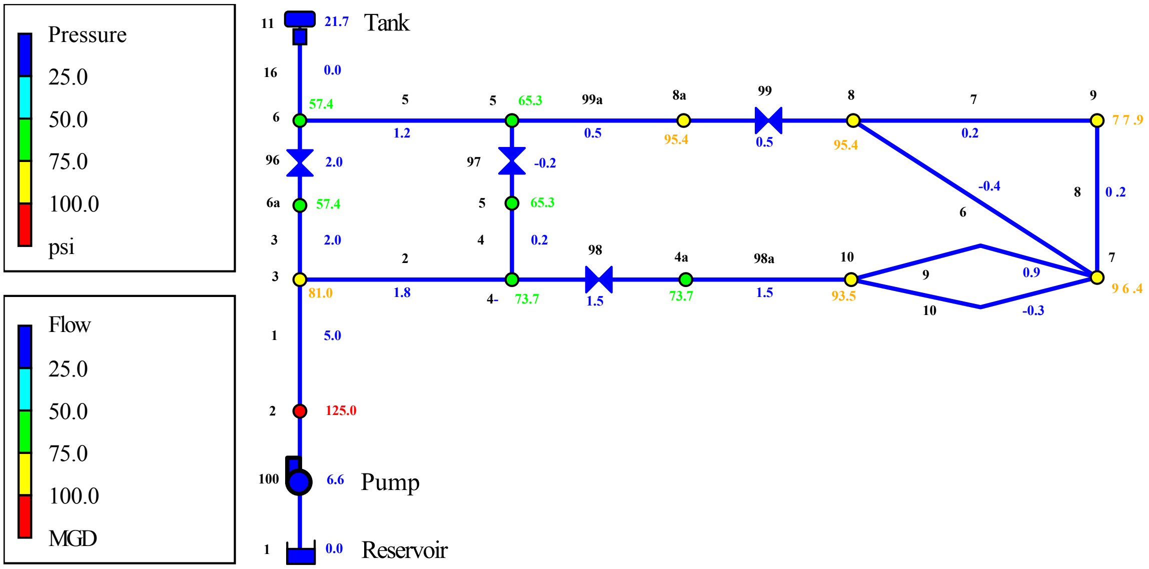


Figure 4.22. Physical infrastructure of the IWDN studied.

This water network is being monitored and controlled by a rule-based system that is considered as the cyber infrastructure. This decision support entity uses data collected from the water system and sends control commands to the actuators with the objective of maintaining the flow of water to the demand nodes.

Table 4.8. Parameters of the IWDN simulated.

Node	Elevation (ft.)	Demand (mgd)	Normal head (ft.)	Minimum head
1	100	-6.62	100	-
2	100	0.73	388.48	146
3	200	1.2	386.43	246
4	210	0.6	376.80	256
5	230	0.4	377.54	276
6	250	0.82	380.05	296
7	10	0.6	173.57	56
8	10	0.8	170.31	56
9	50	0.4	160.87	96
10	25	0.2	181.37	96
8a	10	0	-	-
4a	210	0	-	-
6a	250	0	-	-
5a	230	0	-	-

Pipe	From node / To node	Length (ft.)	Diameter (in.)	Roughness
1	2/3	200	16	120
2	3/4	1500	12	120
3	3/6a	1800	14	120
4	4/5a	2000	10	120
5	6/5	1900	14	120
6	8/7	1000	8	120
7	8/9	2500	10	120
8	7/9	3500	8	120
9	10/7	1500	10	120
10	7/10	1500	6	120
11	11/6	1000	12	100
98a	4a/10	500	6	65
99a	5/8a	500	4	65

4.6.1.2. Definition of system failure. We defined a system failure as comprising one or more of the following three cases:

- Negative pressure: Having a negative pressure in any pipe or at any node
- Water shortage: Having a node which is supplied with less than 80% of its demand
- Excessive outages: Concurrent failure of more than three components

Excessive outages is not a general criterion, but is used to eliminate the need for exhaustive simulation of the entire system states. The number of failures that comprise excessive outage can be set based on the size and number of components of the system.

4.6.1.3. Simulations. An integrated simulation environment is used to understand the operation of water distribution network and to determine the failed and operational states. For the physical infrastructure, we have used EPANET, which is a water distribution network simulator developed by the US Environmental Protection Agency to study functional aspects of the system, such as demand patterns, water quantity (flow and pressure head), and water quality (contaminants and minerals). As EPANET is incapable of simulating intelligent decision support, we constructed specific libraries in MATLAB to simulate the cyber infrastructure. This integrated cyber-physical simulation environment is based on the work presented in [97]. The simulation procedure, depicted in Figure 4.23, was conducted as follows:

1. Set fault conditions
2. Run EPANET and generate operation report
3. Parse the report and extract input for the decision support
4. Operate the decision support algorithm to determine controller settings
5. Output control settings as an EPANET INP file
6. Provide the INP file to EPANET and observe simulation results

EPANET produces a report of the flow and pressure at each node and in each component, as well as a negative pressure warning. These reports were parsed and loaded into MATLAB to determine whether the injected fault conditions result in a system-level failure.

4.6.1.4. Reliability model. As mentioned in Section 4.6.1.2, all triple-component failures are considered system-level failure states. The results of single-component and double-component fault injection were used to generate the reliability model. A number of representative failure scenarios are presented in Table 4.9.

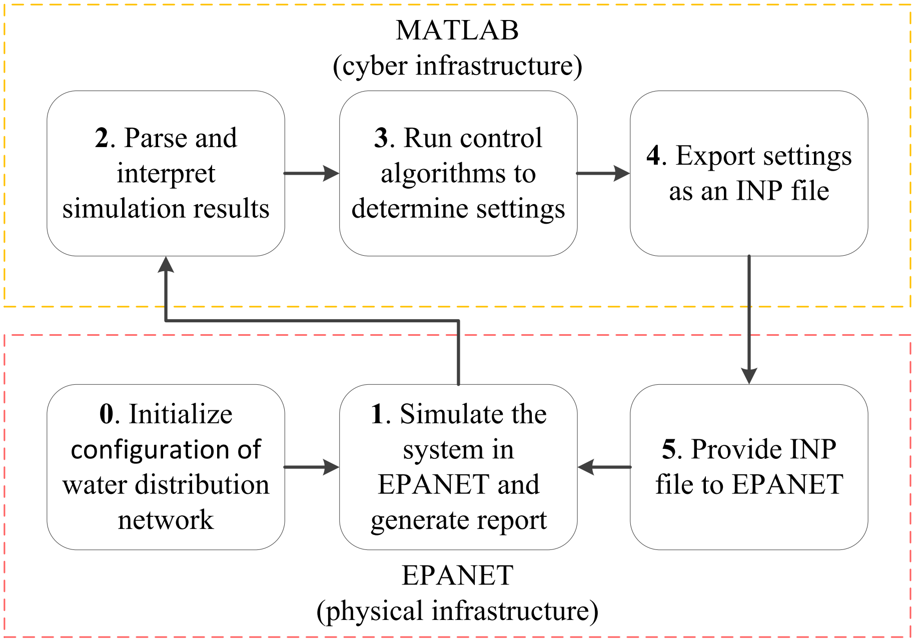

MATLAB
(cyber infrastructure)
2. Parse and interpret simulation results
3. Run control algorithms to determine settings
4. Export settings as an INP file
0. Initialize configuration of water distribution network
1. Simulate the system in EPANET and generate report
5. Provide INP file to EPANET
EPANET
(physical infrastructure)

Figure 4.23. Cyber-physical simulation procedure.

Table 4.9. Representative failure scenarios.

Failed component	State	Failure time	Description
Pump	failed	3	negative pressure at nodes 2, 3, and 4
Tank	operational	-	
Pipe 1	failed	6	negative pressure at node 3 and 4
Pipe 6	operational	-	
Pipe 98a	failed	0	negative pressure
Pipes 7 and 8	failed	0	negative pressure at node 9
Valve 98	failed	3	negative pressure
Valve 99	operational	-	

After identifying the failed and operational states, we can populate the reliability model using the Markov chain imbeddable structure as explained in Section 3.2. Figure 4.24 plots the overall reliability of the system in terms of the reliability of the pipes with three different assumptions about the valves. p_{PI}, p_V, p_T, p_{PU}, and p_{Cyber} respectively, denote reliabilities of each pipe, each valve, each tank, each pump, and the overall reliability of the cyber infrastructure. As expected, we see that improving the reliability of pipes (as the main physical components of a water distribution network) has a significant impact on the reliability of the system. It is also shown that the valves are required to be highly reliable to attain a system with acceptable service, even if the pipes are near perfect. Mathematical representation of the reliability model for this IWDN is shown in Appendix D.

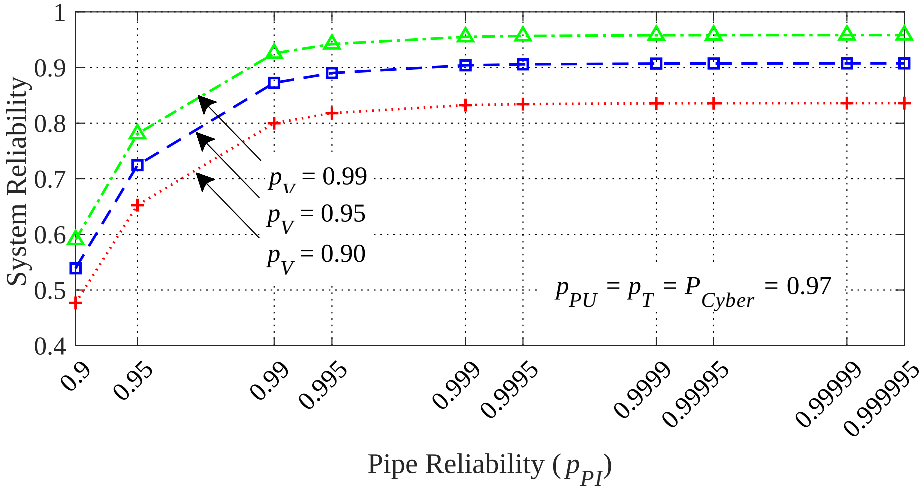


Figure 4.24. Effect of the valves on system-level reliability.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The objective of the research presented in this dissertation is to develop quantitative models for dependability attributes of CPSs. The scope of this work comprises reliability and survivability, as meaningful dependability attributes for analyzing domains of our interest. We have investigated interdependency to enable composition of a unified dependability model from these disparate attributes. Our main focus is on smart grids, as prominent examples of critical infrastructure CPSs. In our case study, we have used the IEEE 14- and 57-bus test systems as bases for developing smart grid examples. To demonstrate generality of our approach, we illustrate the application of our reliability modeling technique to an IWDN.

We have investigated the use of correlation and causation metrics for detection and quantification of the extent of dependency links among the components of a CPS. We have created quantitative interdependency metrics, which seek to capture the effect of multi-step dependencies as well as immediate dependency links. These interdependency metrics reveal important, but previously indiscernible, links among the components. The importance of this revelation is amplified by the fact that some components thus revealed to be strongly dependent are not within geographical, logical, physical, or cyber proximity of each other. This nonlocal property of fault propagation has been observed in the past and was demonstrated through the test cases in this work.

Our reliability analysis reiterates the urgency of improving the computational and communication technologies that underpin modern critical infrastructures. We simulated several failure cases to populate a Markovian reliability model and observed through quantitative analysis that introducing additional interdependency can exponentially degrade system reliability. Propagation of failure from cyber to physical components was found to compromise system-level reliability to the greatest extent.

To better quantify the success of CPSs in achieving their stated goals, such as the ability to autonomously defend against attacks and remediate the consequences of failure, we presented an approach for evaluation of survivability. This approach quantifies attributes pivotal to survivability by determining the rate and extent of degradation of a domain-specific figure-of-merit during a number of selected failure cases. These results were used to identify critical components whose hardening would be most beneficial to survivability.

Our final contribution is a neural networks approach for prediction of imminent component failures. This neural network exhibited excellent predictive performance, which could be attributed in part to the high level of interdependence among components of the systems analyzed. Several related studies from the literature have confirmed the high level of interdependency and corroborated the existence of recurrent failure sequences in most critical infrastructures. These observations support the promise of our failure prediction approach for being efficiently applied to domains other than power.

Proposed avenues for future extension of this research include the following:

- To date, we have assumed that the communication infrastructure will remain functional despite other failures in the critical infrastructure being examined. Considering communication impairments, and data corruption as a manifestation thereof, will refine and increase the accuracy of our dependability model. Corrupted data may originate in malicious attacks to the communication infrastructure or accidental faults during measurement, communication, processing, and storage. While consideration of additional components, e.g., communication links, improves the accuracy of a dependability model, it also increases the size of the system state space and computational complexity. The computational complexity of our proposed modeling approach is not prohibitive for systems with fewer than 500 components; however, a more judicious state elimination method is needed for application to larger-scale CPSs or more refined analysis that examines the system at a higher level of granularity. In systems with independent components, superposition can alleviate the computational burden.

- A challenging task in dependability modeling is to identify essential services of the system being investigated. Equally important is finding an appropriate definition of system-level failure. Qualitative dependability studies, and specifically an ontology, can be utilized to this end. It is very common for a large-scale system to have multiple objectives and provide more than one service. To address this challenge, individual FoMs can be combined to create a multi-dimensional FoM, leading to another avenue for extending our current research.

- Our dependability modeling approaches are useful regardless of whether the disruptive event is caused by an accidental failure or a malicious attack; however, targeted cyber-physical attacks may render plausible a number of specific degraded states omitted from our current analysis. Any omission of plausible degraded states can cause potential overestimation of dependability attributes. This concern inspires more careful consideration of the consequences of cyber attacks in future extensions to our work.

- Tools and techniques for machine learning are ubiquitous and fast-growing, and improving in reliability and accuracy. This allows them to be utilized in critical applications. We have demonstrated applicability of a well-known machine learning tool, the ANN, for prediction of failures in CPSs. As an improvement to this research, state-of-the-art methods such as recurrent neural networks (a more advanced ANN architecture) may be used to incorporate temporal features of the failures as well.

APPENDIX A

NOTABLE INTERDEPENDENCY METRICS FOR IEEE TEST SYSTEMS

In interdependency analysis of the IEEE-14 and IEEE-57 smart grids, weights are assigned to each node as well as each dependency link. In Figure 4.7 and Figure 4.8, these weights are displayed using a graph representation. In this section, numerical values of the weights for notable nodes and links are provided. For easier comparison, we also include the values attained from correlation analysis using the PCC metric. Tables A.1 and A.2 show the top ten direct dependency links for IEEE-14 and IEEE-57 smart grids, respectively. Similarly, Tables A.3 and A.4 indicate the most remarkable pairs of components that despite weak direct dependencies, have large multi-step dependency links. Finally, Tables A.5 through A.8 show the top ten components with largest in-degree/out-degree values for IEEE-14 and IEEE-57 smart grids.

Table A.1. Notable dependencies among components of IEEE-14 smart grid.

PCC		**RDC**		**Causation**	
Components	d_{ij}	Components	d_{ij}	Components	d_{ij}
L_{1-2} - L_{1-5}	0.73	L_{1-2} - L_{1-5}	0.83	L_{1-2} - L_{1-5}	0.83
L_{2-3} - F_{2-3}	0.60	L_{2-3} - F_{2-3}	0.79	L_{2-3} - F_{2-3}	0.81
L_{2-4} - F_{2-4}	0.52	L_{2-4} - F_{2-4}	0.72	L_{2-4} - F_{2-4}	0.77
P_2 - L_{1-2}	0.49	L_{4-5} - L_{1-2}	0.61	L_{1-5} - F_{1-5}	0.56
F_{1-5} - L_{1-2}	0.48	P_2 - L_{1-2}	0.60	F_{1-5} - L_{1-2}	0.39
L_{4-5} - L_{1-2}	0.45	F_{1-5} - L_{1-2}	0.57	P_2 - L_{1-2}	0.32
L_{1-5} - F_{1-5}	0.43	L_{1-5} - F_{1-5}	0.56	L_{1-5} - P_9	0.20
F_{1-5} - L_{1-5}	0.35	F_{1-5} - L_{1-5}	0.52	L_{2-3} - L_{4-5}	0.13
L_{2-3} - L_{1-5}	0.34	L_{1-5} - L_{1-2}	0.46	DS - L_{1-2}	0.12
L_{4-9} - L_{1-2}	0.33	L_{2-3} - L_{1-5}	0.46	L_{7-9} - L_{9-10}	0.12

Table A.2. Notable dependencies among components of IEEE-57 smart grid.

PCC		**RDC**		**Causation**	
Components	d_{ij}	Components	d_{ij}	Components	d_{ij}
L_{1-17} - F_{1-17}	0.76	L_{7-29} - L_{8-9}	0.96	L_{12-17} - F_{12-17}	0.95
L_{7-29} - L_{8-9}	0.73	L_{12-17} - F_{12-17}	0.94	L_{1-17} - F_{1-17}	0.94
L_{12-17} - F_{12-17}	0.72	L_{1-17} - F_{1-17}	0.93	L_{28-29} - P_{28}	0.92
L_{2-3} - L_{1-15}	0.69	L_{28-29} - P_{28}	0.92	L_{7-29} - L_{8-9}	0.92
F_{1-16} - F_{1-16}	0.68	L_{34-35} - P_{32}	0.92	L_{1-16} - F_{1-16}	0.91
L_{7-8} - F_{7-8}	0.67	L_{9-55} - P_{53}	0.92	L_{37-38} - P_{56}	0.91
L_{34-35} - P_{32}	0.65	L_{37-38} - P_{32}	0.92	L_{7-8} - F_{7-8}	0.91
L_{28-29} - P_{28}	0.65	L_{37-38} - P_{56}	0.92	L_{6-7} - F_{6-7}	0.90
L_{9-55} - P_{53}	0.65	L_{34-32} - P_{32}	0.91	L_{54-55} - F_{54-55}	0.88
L_{37-38} - P_{32}	0.65	L_{35-36} - P_{32}	0.91	L_{7-29} - P_{28}	0.87

Table A.3. Notable multi-step dependency links among components of IEEE-14 smart grid; dependency links that are relatively large in **T**, but small in **D**.

PCC	**RDC**	**Causation**
Components	Components	Components
L_{2-3} - L_{1-2}	L_{2-3} - L_{1-2}	L_{1-2} - F_{1-5}
L_{4-5} - L_{1-5}	L_{2-3} - L_{1-5}	F_{1-5} - L_{1-5}
L_{2-3} - L_{1-5}	L_{4-5} - L_{1-5}	P_{2} - L_{1-5}
L_{2-4} - L_{1-5}	L_{2-4} - L_{1-5}	L_{1-5} - L_{1-2}
L_{2-3} - F_{1-5}	L_{2-3} - F_{1-5}	L_{1-2} - P_{9}
L_{2-3} - L_{1-2}	L_{2-4} - L_{1-2}	P_{2} - F_{1-5}
P_{2} - L_{1-5}	L_{2-4} - F_{1-5}	DS - L_{1-5}
L_{2-4} - L_{1-2}	L_{4-5} - F_{1-5}	F_{1-5} - P_{9}
L_{2-4} - F_{1-5}	L_{5-6} - L_{1-5}	L_{4-5} - L_{1-5}
L_{5-6} - L_{1-5}	L_{7-9} - L_{1-5}	DS - F_{1-5}

Table A.4. Notable multi-step dependency links among components of IEEE-57 smart grid; dependency links that are relatively large in **T**, but small in **D**.

PCC	**RDC**	**Causation**
Components	Components	Components
L_{7-29} - L_{1-15}	L_{7-29} - L_{1-15}	L_{7-29} - L_{7-8}
L_{7-29} - L_{1-2}	L_{7-29} - L_{1-2}	L_{7-8} - F_{6-8}
L_{7-29} - F_{7-8}	L_{7-29} - F_{7-8}	L_{8-9} - F_{7-8}
L_{7-29} - P_{32}	L_{1-16} - P_{32}	L_{1-16} - L_{1-2}
L_{7-29} - F_{1-17}	L_{7-29} - F_{1-17}	L_{7-29} - L_{6-8}
L_{7-29} - L_{1-17}	L_{7-29} - L_{1-17}	L_{1-15} - F_{1-17}
L_{1-16} - P_{32}	L_{7-29} - P_{32}	L_{8-9} - F_{6-8}
L_{7-29} - L_{7-8}	L_{1-2} - P_{32}	L_{1-2} - F_{1-17}
L_{1-15} - P_{32}	L_{1-15} - P_{32}	P_{32} - F_{1-17}
L_{8-9} - L_{1-15}	L_{1-16} - F_{1-17}	L_{1-2} - F_{1-16}

Table A.5. Largest weighted out-degree values in the IEEE-14 smart grid.

PCC		**RDC**		**Causation**	
Component	τ	Components	τ	Component	τ
L_{2-3}	0.239	L_{2-3}	0.244	L_{1-2}	0.206
L_{4-5}	0.219	L_{4-5}	0.234	L_{1-5}	0.187
L_{2-4}	0.215	L_{2-4}	0.220	L_{4-5}	0.152
L_{1-2}	0.205	L_{1-2}	0.206	F_{1-5}	0.143
P_9	0.196	L_{7-9}	0.200	L_{2-3}	0.128
L_{1-5}	0.193	P_2	0.199	P_2	0.127
F_{1-5}	0.188	L_{5-6}	0.197	L_{2-4}	0.113
P_2	0.187	L_{1-5}	0.196	P_9	0.093
F_{2-3}	0.185	F_{1-5}	0.194	DS	0.091
L_{5-6}	0.178	P_9	0.191	L_{7-9}	0.071

Table A.6. Largest weighted out-degree values in the IEEE-57 smart grid.

PCC		**RDC**		**Causation**	
Component	τ	Components	τ	Component	τ
L_{7-29}	0.073	L_{7-29}	0.072	L_{1-2}	0.047
L_{8-9}	0.069	L_{1-16}	0.067	L_{7-29}	0.038
L_{1-2}	0.063	L_{1-2}	0.067	L_{1-15}	0.033
L_{1-16}	0.063	L_{8-9}	0.067	L_{1-16}	0.031
L_{1-15}	0.062	L_{1-15}	0.063	L_{7-8}	0.028
L_{7-8}	0.055	L_{3-4}	0.060	L_{22-23}	0.025
P_{25}	0.053	L_{22-23}	0.060	L_{8-9}	0.024
L_{12-17}	0.052	L_{37-39}	0.058	L_{1-17}	0.018
P_{19}	0.051	L_{7-8}	0.056	L_{37-38}	0.018
P_{28}	0.050	L_{12-17}	0.054	L_{6-8}	0.016

Table A.7. Largest weighted in-degree values in the IEEE-14 smart grid.

PCC		**RDC**		**Causation**	
Component	ν	Components	ν	Component	ν
L_{1-2}	0.591	L_{1-5}	0.604	L_{1-5}	0.181
L_{1-5}	0.575	L_{1-2}	0.594	L_{1-2}	0.1168
F_{1-5}	0.497	F_{1-5}	0.495	F_{1-5}	0.151
P_9	0.421	P_9	0.427	P_9	0.148
F_{2-3}	0.318	F_{2-3}	0.322	L_{4-5}	0.118
F_{2-4}	0.269	L_{4-5}	0.285	L_{9-10}	0.103
L_{4-5}	0.263	F_{2-4}	0.280	F_{2-3}	0.101
L_{2-3}	0.219	L_{9-10}	0.237	L_{7-9}	0.093
L_{9-10}	0.187	L_{2-3}	0.210	F_{2-4}	0.088
L_{7-9}	0.175	L_{7-9}	0.208	L_{2-3}	0.080

Table A.8. Largest weighted in-degree values in the IEEE-57 smart grid.

PCC		**RDC**		**Causation**	
Component	v	Components	v	Component	v
F_{1-17}	0.130	L_{1-17}	0.132	P_{32}	0.045
L_{1-17}	0.130	F_{1-17}	0.128	L_{1-17}	0.043
P_{32}	0.116	P_{32}	0.126	F_{1-17}	0.031
L_{1-15}	0.101	L_{1-15}	0.109	L_{1-15}	0.028
L_{1-2}	0.086	L_{1-2}	0.099	P_{28}	0.027
P_{28}	0.083	P_{25}	0.092	P_{53}	0.023
P_{25}	0.081	P_{28}	0.092	L_{8-9}	0.022
P_{53}	0.081	P_{53}	0.089	P_{25}	0.020
F_{1-16}	0.077	F_{1-16}	0.084	L_{7-8}	0.019
F_{7-8}	0.077	P_{56}	0.084	P_{56}	0.019

APPENDIX B

ANN FOR DYNAMIC TUNING OF SSSC DEVICES

For each of the IEEE-14 and IEEE-57 smart grid systems, we designed and trained an ANN to provide decision support. Each of these ANNs has one input layer, fed with parameters describing system status (i.e., bus voltages and line power flows), one hidden layer, and one output layer, providing settings for SSSC devices. Each ANN is trained by a lookup table generated by exhaustive search for optimal settings on N-1 failure cases. The exhaustive search is performed seeking for the settings that minimize line overloads and is calculated as shown in Equation (B.1).

$$\theta = \sum_{i=1}^{n} \left(\frac{P_i}{C_i} \right)^{2\mu} \tag{B.1}$$

where n is the number of lines; P_i and C_i are the active power flow and maximum capacity of line i, respectively; μ controls the extent to which a given setting should be penalized for line overloads. We have set $\mu = 5$.

Figure B.1 shows the architecture of the ANN used for dynamic adjustment of SSSC settings.

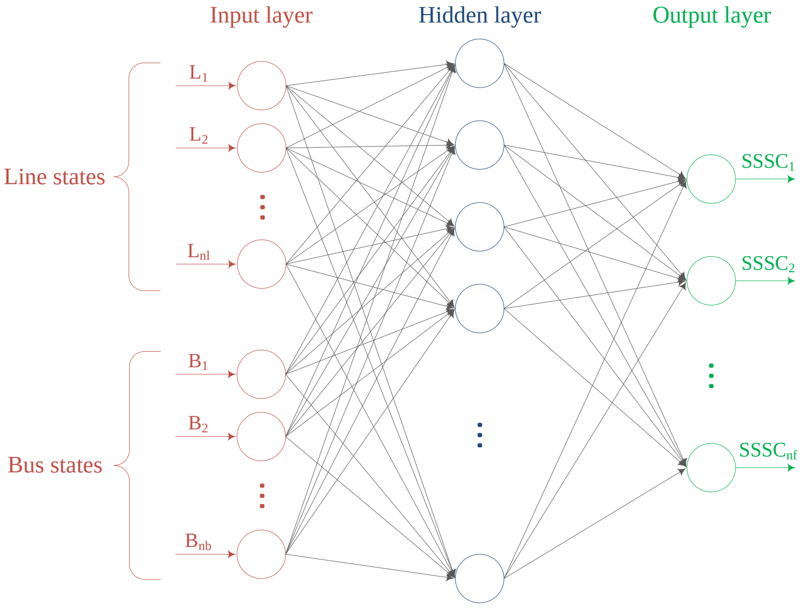


Figure B.1. Architecture of the ANN used for decision support of IEEE test cases.

Table B.1 shows for each of the IEEE test cases, the number of nodes at each layer.

Table B.1. Number of nodes at each layer of the ANN used as decision support for IEEE test cases.

	IEEE-14	IEEE-57
input layer	34	137
hidden layer	20	75
output layer	3	7

The R^2 measure of goodness of fit for the ANNs trained for IEEE-14 and IEEE-57 smart grids are 0.92 and 0.90, respectively.

APPENDIX C

RELIABILITY OF IEEE 14-BUS SMART GRID

System-level reliability of IEEE-14 smart grid in fail-bypass mode is shown in Equation (C.1), where p_{DS}, p_F, p_P, p_L, and p_{CM}, respectively, represent the (component-level) reliability of the control algorithm, each SSSC device, each PMU device, each transmission line, and each communication link. For every component-level reliability value p, the corresponding unreliability is defined as $q = 1 - p$. For a better tractability, but without loss of generality, all components of the same type (e.g., all PMUs) are considered to be equally reliable.

$$\begin{aligned}
R_{sys} = p_L^{20} +& \\
19 p_{DS}.{p_F}^3.{p_P}^4.q_L.{p_L}^{19}.{p_{CM}}^7 +& \\
167 p_{DS}.{p_F}^3.{p_P}^4.{q_L}^2.{p_L}^{18}.{p_{CM}}^7 +& \\
76 p_{DS}.{p_F}^3.q_P.{p_P}^3.q_L.{p_L}^{19}.{p_{CM}}^6 +& \\
76 p_{DS}.{p_F}^3.{p_P}^4.q_L.{p_L}^{19}.q_{CM}.{p_{CM}}^6 +& \\
668 p_{DS}.{p_F}^3.q_P.{p_P}^3.{q_L}^2.{p_L}^{18}.{p_{CM}}^6 +& \\
835 p_{DS}.{p_F}^3.{p_P}^4.{q_L}^2.{p_L}^{18}.q_{CM}.{p_{CM}}^6 +& \\
56 p_{DS}.q_F.{p_F}^2.{p_P}^4.q_L.{p_L}^{19}.{p_{CM}}^6 +& \\
481 p_{DS}.q_F.{p_F}^2.{p_P}^4.{q_L}^2.{p_L}^{18}.{p_{CM}}^6 +& \\
219 p_{DS}.q_F.{p_F}^2.q_P.{p_P}^3.q_L.{p_L}^{19}.{p_{CM}}^5 +& \\
219 p_{DS}.q_F.{p_F}^2.{p_P}^4.q_L.{p_L}^{19}.q_{CM}.{p_{CM}}^5 +& \\
1849 p_{DS}.q_F.{p_F}^2.q_P.{p_P}^3.{q_L}^2.{p_L}^{18}.{p_{CM}}^5 +& \\
1849 p_{DS}.q_F.{p_F}^2.{p_P}^4.{q_L}^2.{p_L}^{18}.q_{CM}.{p_{CM}}^5 +& \\
16 q_{DS}.{p_F}^3.q_L.{p_L}^{19} +& \\
393 q_{DS}.{p_F}^3.{q_L}^2.{p_L}^{18} +& \\
221 q_{DS}.q_F.{p_F}^2.q_L.{p_L}^{19} +& \\
1705 q_{DS}.q_F.{p_F}^2.{q_L}^2.{p_L}^{18}&
\end{aligned} \tag{C.1}$$

Equation (C.1) can be interpreted as the sum of probabilities of being in any of the states that do not lead to a failure (as defined in Section 4.3.1). The system can withstand cyber impairments in virtue of the relatively conservative fail-bypass mode, as manifested by the term p_L^{20}. The subsequent terms correspond to the cases where at least one transmission lines is in outage, and hence, fault-free operation of the cyber network is needed in order to mitigate the impacts of disruptions imposed to the system.

APPENDIX D

RELIABILITY OF INTELLIGENT WATER DISTRIBUTION NETWORK

System-level reliability of the IWDN studied in this document is as shown in Equation (D.1), where p_{PU}, p_T, p_{PI}, p_V, and p_{Cyber}, respectively, represent the component-level reliability of the pump, the tank, each pipe, each valve, and the overall reliability of the cyber infrastructure. For each component-level reliability value, p, the corresponding unreliability is defined as $q = 1 - p$. For better tractability, but without loss of generality, all components of the same type (e.g., all pipes) are considered to be equally reliable.

$$
\begin{aligned}
R_{sys} = {} & p_{PU}.p_T.{p_{PI}}^{13}.{p_V}^{4}.p_{Cyber} + \\
& p_{PU}.q_T.{p_{PI}}^{13}.{p_V}^{4}.p_{Cyber} + \\
& 8p_{PU}.q_T.{p_{PI}}^{12}.q_{PI}.{p_V}^{4}.p_{Cyber} + \\
& p_{PU}.q_T.{p_{PI}}^{13}.{p_V}^{4}.q_{Cyber} + \\
& 10p_{PU}.p_T.{p_{PI}}^{12}.q_{PI}.{p_V}^{4}.p_{Cyber} + \\
& 31p_{PU}.p_T.{p_{PI}}^{11}.{q_{PI}}^{2}.{p_V}^{4}.p_{Cyber} + \\
& 16p_{PU}.p_T.{p_{PI}}^{12}.q_{PI}.{p_V}^{3}.q_V.p_{Cyber} + \\
& 5p_{PU}.p_T.{p_{PI}}^{12}.q_{PI}.{p_V}^{4}.q_{Cyber} + \\
& 3p_{PU}.p_T.{p_{PI}}^{13}.{p_V}^{3}.q_V.p_{Cyber} + \\
& p_{PU}.p_T.{p_L}^{13}.{p_V}^{4}.q_{Cyber}
\end{aligned}
\tag{D.1}
$$

Equation (D.1) can be interpreted as the sum of probabilities of being in any of the states that do not lead to a failure (as defined in Section 4.6.1.2).

REFERENCES

[1] "NIST framework and roadmap for smart grid interoperability standards, release 1.0," tech. rep., National Institute of Standards and Technology, January 2010.

[2] "Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," tech. rep., U.S.-Canada Power System Outage Task Force, April 2004. Retrieved 2017-05-02.

[3] A. Berizzi, "The Italian 2003 blackout," in *IEEE Power Engineering Society General Meeting*, pp. 1673–1679, 2004.

[4] "Hurricane Sandy rebuilding strategy," tech. rep., Hurricane Sandy Rebuilding Task Force, August 2013. Retrieved 2017-05-02.

[5] S. Miles, H. Gallagher, and C. Huxford, "Restoration and impacts from the September 8, 2011 San Diego power outage," *Journal of Infrastructure Systems*, vol. 19, no. 3, 2013.

[6] Federal Energy Regulatory Commission and North American Electric Reliability Corporation, "Arizona-Southern California outages on September 8, 2011-causes and recommendations," April 2012. Retrieved 2017-05-02.

[7] P. Derler, E. Lee, and A. Vincentelli, "Modeling cyber-physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2012.

[8] K. Marashi and S. Sedigh Sarvestani, "Towards comprehensive modeling of reliability for smart grids: Requirements and challenges," in *Proceedings of the 15th IEEE International High Assurance Systems Engineering Symposium (HASE)*, (Miami, FL), pp. 105–112, January 2014.

[9] K. Marashi, M. Woodard, S. Sedigh Sarvestani, and A. R. Hurson, "Quantitative reliability analysis for intelligent water distribution networks," in *Proceedings of the Embedded Topical Meeting on Risk Management for Complex Socio-Technical Systems (RM4CSS), Annual Meeting of the American Nuclear Society*, (Washington, D.C.), November 2013.

[10] K. Marashi, M. Woodard, S. Sedigh Sarvestani, and A. R. Hurson, "Quantitative reliability analysis for intelligent water distribution networks," in *Risk Management for Complex Socio-Technical Systems*, American Nuclear Society, to appear.

[11] K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, "Consideration of cyber-physical interdependencies in reliability modeling of smart grids," *IEEE Transactions on Sustainable Computing – Special Issue on Sustainable Cyber-Physical Systems*, to appear.

[12] M. Woodard, K. Marashi, and S. Sedigh Sarvestani, "Survivability evaluation and importance analysis for complex networked systems," *IEEE Transactions on Network Science and Engineering*, under review.

[13] K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, "Identification of interdependencies and prediction of fault propagation for cyber-physical systems," *Reliability Engineering & System Safety*, to be submitted.

[14] K. Marashi, S. Sedigh Sarvestani, and A. R. Hurson, "Quantification and analysis of interdependency in cyber-physical systems," in *Proceedings of of the 3rd International Workshop on Reliability and Security Aspects for Critical Infrastructure (ReSA4CI 2016), in conjunction with the 46th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2016)*, (Toulouse, France), pp. 149–154, June 2016.

[15] N. Jarus, M. Woodard, K. Marashi, A. Faza, J. Lin, P. Maheshwari, and S. Sedigh Sarvestani, "Survey on modeling and design of cyber-physical systems," *ACM Transactions on Cyber-Physical Systems*, under review.

[16] "Grid 2030: A national vision for electricity's second 100 years," tech. rep., United States Department of Energy Office of Electric Transmission and Distribution, 2003. Retrieved 2017-05-02.

[17] United States Environmental Protection Agency, "EPANET2 User's manual." http://www.innovyze.com/products/epanet/download/P1007WWU.pdf. Retrieved 2017-05-02.

[18] F. Milano, "An open source power system analysis toolbox," *IEEE Transactions on Power Systems*, vol. 20, pp. 1199–1206, August 2005.

[19] W. Kuo and M. Zuo, *Optimal Reliability Modeling: Principles and Applications*. Wiley, 2003.

[20] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, pp. 11–33, January 2004.

[21] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, and T. Longstaff, "Survivable network systems: An emerging discipline," tech. rep., DTIC Document, 1997. Retrieved 2017-05-02.

[22] J. F. Meyer, "On evaluating the performability of degradable computing systems," *IEEE Transactions on Computers*, vol. 100, no. 8, pp. 720–731, 1980.

[23] D. Henry and J. E. Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114 – 122, 2012.

[24] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez, “A review of definitions and measures of system resilience,” *Reliability Engineering & System Safety*, vol. 145, pp. 47 – 61, 2016.

[25] R. Ball, *The fundamentals of aircraft combat survivability analysis and design*. No. v. 1 in AIAA education series, American Institute of Aeronautics and Astronautics, 2003.

[26] National Communications System, Technology & Standards Division, “Telecommunications: Glossary of telecommunication terms - federal standard 1037C.” http://www.its.bldrdoc.gov/fs-1037/fs-1037c.htm, 1996. Retrieved 2017-05-02.

[27] M. S. Deutsch and R. R. Willis, *Software Quality Engineering: A Total Technical and Management Approach*. Software Engineering, Prentice Hall, 1988.

[28] ISO/IEC/IEEE, “Systems and software engineering – vocabulary,” *ISO/IEC/IEEE 24765:2010(E)*, pp. 1–418, December 2010.

[29] A. Avritzer, L. Carnevali, L. Happe, B. R. Haverkort, A. Koziolek, D. Menasché, A. Remke, and S. Sedigh Sarvestani, “Survivability evaluation of gas, water and electricity infrastructures,” in *Proceedings of the 7th International Workshop on Practical Applications of Stochastic Modelling (PASM)*, (Newcastle upon Tyne, UK), May 2015.

[30] P. E. Heegaard and K. S. Trivedi, “Network survivability modeling,” *Computer Networks*, vol. 53, no. 8, pp. 1215 – 1234, 2009.

[31] T1A1.2 Working Group, “Technical report on enhanced network survivability performance,” tech. rep., T1A1.2 Working Group on Network Survivability Performance, February 2001.

[32] Z. S. Ma, “A unified definition for reliability, survivability and resilience inspired by the handicap principle and ecological stability,” *International Journal of Critical Infrastructures*, vol. 8, no. 2-3, pp. 242–272, 2012.

[33] D. S. Menasché, R. M. Meri Leao, E. de Souza e Silva, A. Avritzer, S. Suresh, K. Trivedi, R. A. Marie, L. Happe, and A. Koziolek, “Survivability analysis of power distribution in smart grids with active and reactive power modeling,” *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, pp. 53–57, January 2012.

[34] P. Chopade and M. Bikdash, “Modeling for survivability of smart power grid when subject to severe emergencies and vulnerability,” in *Proceedings of IEEE Southeastcon*, pp. 1–6, March 2012.

[35] I. A. Alobaidi, S. Sedigh Sarvestani, and A. R. Hurson, “Survivability analysis and recovery support for smart grids,” in *Proceedings of the 4th International Symposium on Resilient Cyber Systems*, (Chicago, USA), August 2016.

[36] J. C. Knight and K. J. Sullivan, “On the definition of survivability,” *University of Virginia, Department of Computer Science, Technical Report CS-TR-33-00*, 2000.

[37] A. Volkanovski, M. Cepin, and B. Mavko, "Application of the fault tree analysis for assessment of power system reliability," *Reliability Engineering & System Safety*, vol. 94, no. 6, pp. 1116 – 1127, 2009.

[38] E. Zio and L. Golea, "Analyzing the topological, electrical and reliability characteristics of a power transmission system for identifying its critical elements," *Reliability Engineering & System Safety*, vol. 101, pp. 67 – 74, May 2012.

[39] C. Singh and A. Lago-Gonzalez, "Reliability modeling of generation systems including unconventional energy sources," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, pp. 1049–1056, May 1985.

[40] G. Celli, E. Ghiani, F. Pilo, and G. G. Soma, "Reliability assessment in smart distribution networks," *Electric Power Systems Research*, vol. 104, pp. 164 – 175, 2013.

[41] P. Venemans and M. Schreuder, "A method for the quantitative assessment of reliability of smart grids," in *CIRED Workshop, Integration of Renewables into the Distribution Grid*, pp. 1–4, May 2012.

[42] B. Falahati, Y. Fu, and L. Wu, "Reliability assessment of smart grid considering direct cyber-power interdependencies," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1515–1524, September 2012.

[43] B. Falahati and Y. Fu, "Reliability assessment of smart grids considering indirect cyber-power interdependencies," *IEEE Transactions on Smart Grid*, vol. 5, pp. 1677–1685, July 2014.

[44] K. Moslehi and R. Kumar, "A reliability perspective of the smart grid," *IEEE Transactions on Smart Grid*, vol. 1, pp. 57–64, June 2010.

[45] A. Dominguez-Garcia, "Reliability modeling of cyber-physical electric power systems: A system-theoretic framework," in *IEEE Power and Energy Society General Meeting*, (San Diego, CA), pp. 1–5, July 2012.

[46] P. Gross, A. Boulanger, M. Arias, D. Waltz, P. M. Long, C. Lawson, R. Anderson, M. Koenig, M. Mastrocinque, W. Fairechio, J. A. Johnson, S. Lee, F. Doherty, and A. Kressner, "Predicting electricity distribution feeder failures using machine learning susceptibility analysis," in *Proceedings of the 18th Conference on Innovative Applications of Artificial Intelligence*, vol. 2, pp. 1705–1711, 2006.

[47] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. N. Gross, B. Huang, S. Ierome, D. F. Isaac, A. Kressner, R. J. Passonneau, A. Radeva, and L. Wu, "Machine learning for the New York City power grid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 328–345, February 2012.

[48] J.-C. Laprie, K. Kanoun, and M. Kaâniche, “Modelling interdependencies between the electricity and information infrastructures,” in *Proceedings of the 26th International Conference on Computer Safety, Reliability and Security (SAFECOMP)*, vol. 4680, pp. 54–67, 2007.

[49] M. Beccuti, S. Chiaradonna, F. D. Giandomenico, S. Donatelli, G. Dondossola, and G. Franceschinis, “Quantification of dependencies between electrical and information infrastructures,” *International Journal of Critical Infrastructure Protection*, vol. 5, pp. 14 – 27, March 2012.

[50] E. Casalicchio and E. Galli, “Metrics for quantifying interdependencies,” in *Critical Infrastructure Protection II*, vol. 290, pp. 215–227, 2008.

[51] Z. Huang, C. Wang, S. Ruj, M. Stojmenovic, and A. Nayak, “Modeling cascading failures in smart power grid using interdependent complex networks and percolation theory,” in *Proceedings of the 8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1023–1028, June 2013.

[52] Z. Huang, C. Wang, M. Stojmenovic, and A. Nayak, “Characterization of cascading failures in interdependent cyber-physical systems,” *IEEE Transactions on Computers*, vol. 64, pp. 2158–2168, August 2015.

[53] D. Zhou, J. Gao, H. E. Stanley, and S. Havlin, “Percolation of partially interdependent scale-free networks,” *Physical Review E*, vol. 87, p. 052812, May 2013.

[54] A. D. Giorgio and F. Liberati, “A Bayesian network-based approach to the critical infrastructure interdependencies analysis,” *IEEE Systems Journal*, vol. 6, pp. 510–519, September 2012.

[55] D. Mendonça and W. A. Wallace, “Impacts of the 2001 World Trade Center attack on New York City critical infrastructures,” *Journal of Infrastructure Systems*, vol. 12, pp. 260–270, December 2006.

[56] L. Duenas-Osorio and A. Kwasinski, “Quantification of lifeline system interdependencies after the 27 February 2010 Mw 8.8 offshore Maule, Chile, earthquake,” *Earthquake Spectra*, vol. 28, no. S1, pp. S581–S603, 2012.

[57] R. Filippini and A. Silva, “A modeling framework for the resilience analysis of networked systems-of-systems based on functional dependencies,” *Reliability Engineering & System Safety*, vol. 125, pp. 82 – 91, 2014.

[58] E. Bompard, R. Napoli, and F. Xue, “Assessment of information impacts in power system security against malicious attacks in a general framework,” *Reliability Engineering & System Safety*, vol. 94, no. 6, pp. 1087 – 1094, 2009.

[59] S. Rinaldi, J. Peerenboom, and T. Kelly, “Identifying, understanding, and analyzing critical infrastructure interdependencies,” *IEEE Control Systems Magazine*, vol. 11, pp. 11–25, December 2001.

[60] M. Ouyang, “Review on modeling and simulation of interdependent critical infrastructure systems,” *Reliability Engineering & System Safety*, vol. 121, pp. 43 – 60, 2014.

[61] T. Verma, W. Ellens, and R. E. Kooij, “Context-independent centrality measures underestimate the vulnerability of power grids,” *International Journal of Critical Infrastructures*, vol. 11, no. 1, pp. 62–81, 2015.

[62] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, “Do topological models provide good information about electricity infrastructure vulnerability?,” *Chaos*, vol. 20, no. 3, p. 033122, 2010.

[63] S. Asgarpoor and M. Mathine, “Reliability evaluation of distribution systems with non-exponential down times,” *IEEE Transactions on Power Systems*, vol. 12, pp. 579–584, May 1997.

[64] I. Tien and A. D. Kiureghian, “Algorithms for Bayesian network modeling and reliability assessment of infrastructure systems,” *Reliability Engineering & System Safety*, vol. 156, pp. 134 – 147, 2016.

[65] C. Singh and A. Sprintson, “Reliability assurance of cyber-physical power systems,” in *IEEE Power and Energy Society General Meeting*, pp. 1–6, July 2010.

[66] A. Bose, “Models and techniques for the reliability analysis of the smart grid,” in *IEEE Power and Energy Society General Meeting*, pp. 1–5, July 2010.

[67] Z. Zhang, W. An, and F. Shao, “Cascading failures on reliability in cyber-physical system,” *IEEE Transactions on Reliability*, vol. 65, pp. 1745–1754, Dec 2016.

[68] Y. Wang, W. Li, and J. Lu, “Reliability analysis of wide-area measurement system,” *IEEE Transactions on Power Delivery*, vol. 25, pp. 1483–1491, March 2010.

[69] V. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. Hancke, “Smart grid technologies: Communication technologies and standards,” *IEEE Transactions on Industrial Informatics*, vol. 7, pp. 529–539, November 2011.

[70] A. Avritzer, S. Suresh, D. S. Menasché, R. M. M. Leao, E. de Souza e Silva, M. C. Diniz, K. Trivedi, L. Happe, and A. Koziolek, “Survivability models for the assessment of smart grid distribution automation network designs,” in *Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering (ICPE)*, (New York, NY, USA), pp. 241–252, 2013.

[71] A. Koziolek, A. Avritzer, S. Suresh, D. Sadoc Menasché, K. Trivedi, and L. Happe, “Design of distribution automation networks using survivability modeling and power flow equations,” in *IEEE 24th International Symposium on Software Reliability Engineering (ISSRE)*, pp. 41–50, November 2013.

[72] D. S. Menasché, A. Avritzer, S. Suresh, R. M. Leao, E. de Souza e Silva, M. Diniz, K. Trivedi, L. Happe, and A. Koziolek, "Assessing survivability of smart grid distribution network designs accounting for multiple failures," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 12, pp. 1949–1974, 2014.

[73] Z. Yi and T. Dohi, "A simulation approach to quantify network survivability on MANETs," in *Proceedings of the 39th IEEE Computer Software and Applications Conference (COMPSAC)*, vol. 3, pp. 268–273, July 2015.

[74] D. Lopez-Paz, P. Hennig, and B. Schölkopf, "The randomized dependence coefficient," in *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

[75] J. Qi, K. Sun, and S. Mei, "An interaction model for simulation and mitigation of cascading failures," *IEEE Transactions on Power Systems*, vol. 30, pp. 804–819, March 2015.

[76] A. Z. Faza, S. Sedigh Sarvestani, and B. M. McMillin, "The advanced electric power grid: Complexity reduction techniques for reliability modeling," in *Proceedings of the 27th international conference on Computer Safety, Reliability, and Security (SAFECOMP)*, pp. 429–439, 2008.

[77] G. Ou and Y. L. Murphey, "Multi-class pattern classification using neural networks," *Pattern Recognition*, vol. 40, no. 1, pp. 4 – 18, 2007.

[78] C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia, "Incorporating second-order functional knowledge for better option pricing," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, (Cambridge, MA, USA), pp. 451–457, 2000.

[79] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference for Learning Representations*, 2015.

[80] University of Washington, "Power systems test case archive." http://www.ee.washington.edu/research/pstca/. Retrieved 2017-05-02.

[81] Y. Song and B. Wang, "Survey on reliability of power electronic systems," *IEEE Transactions on Power Electronics*, vol. 28, no. 1, pp. 591–604, 2013.

[82] A. Z. Faza, S. Sedigh Sarvestani, and B. M. McMillin, "Reliability analysis for the advanced electric power grid: From cyber control and communication to physical manifestations of failure," in *Proceedings of the 28th International Conference on Computer Safety, Reliability and Security (SAFECOMP)*, vol. 5775, (Hamburg, Germany), pp. 257–269, September 2009.

[83] R. Duffey and T. Ha, "The probability and timing of power system restoration," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 3–9, 2013.

[84] M. Asprou and E. Kyriakides, "Optimal PMU placement for improving hybrid state estimator accuracy," in *IEEE Trondheim PowerTech*, pp. 1–7, June 2011.

[85] N. Acharya and N. Mithulananthan, “Locating series FACTS devices for congestion management in deregulated electricity markets,” *Electric Power Systems Research*, vol. 77, pp. 352–360, March 2007.

[86] T. T. Nguyen and R. Gianto, “Neural networks for adaptive control coordination of PSSs and FACTS devices in multimachine power system,” *IET Generation, Transmission & Distribution*, vol. 2, no. 3, pp. 355–372, 2008.

[87] W. Qiao, R. Harley, and G. Venayagamoorthy, “Neural-network-based intelligent control for improving dynamic performance of facts devices,” in *iREP Symposium - Bulk Power System Dynamics and Control - VII, Revitalizing Operational Reliability*, pp. 1–9, 2007.

[88] “PowerWorld Corporation.” http://www.powerworld.com/products/simulator/overview. Retrieved 2017-05-02.

[89] “DIgSILENT GmbH.” http://www.digsilent.com/. Retrieved 2017-05-02.

[90] “MATPOWER, a MATLAB power system simulation package.” http://www.pserc.cornell.edu/matpower/. Retrieved 2017-05-02.

[91] “EN 50160, voltage characteristics of electricity supplied by public distribution systems,” tech. rep., CENELEC, 2005.

[92] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, “Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines,” *Computer Networks*, vol. 54, pp. 1245–1265, June 2010.

[93] B. Roczen, R. G. Arno, and P. S. Hale, “Reliability block diagram methodology applied to gold book standard network,” in *Proceedings of the IEEE Industrial and Commercial Power Systems Technical Conference*, pp. 116–126, May 2004.

[94] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Network robustness and fragility: Percolation on random graphs,” *Physical Review Letters*, vol. 85, pp. 5468–5471, December 2000.

[95] J. Wagner, U. Shamir, and D. Marks, “Water distribution reliability: Analytical methods,” *Journal of Water Resources Planning and Management*, vol. 114, no. 3, pp. 253–275, 1988.

[96] J. Wagner, U. Shamir, and D. Marks, “Water distribution reliability: Simulation methods,” *Journal of Water Resources Planning and Management*, vol. 114, no. 3, pp. 276–294, 1988.

[97] J. Lin, S. Sedigh Sarvestani, and A. Miller, “Integrated cyber-physical simulation of intelligent water distribution networks,” *Scientic and Engineering Applications Using MATLAB*, 2011.

VITA

Koosha Marashi was born in Isfahan, Iran. He received his B.Sc. in Electrical Engineering from the Isfahan University of Technology in February 2011. He continued his studies in Computer Engineering at the Missouri University of Science and Technology in August 2012. During his Ph.D. studies, Koosha completed three consecutive summer internships at Kalscott Eng. and Intellispeak LLC and gained practical experience in hardware and software development. In July 2017, he received his Ph.D. degree in Computer Engineering from Missouri University of Science and Technology.