Vulnerability Analysis and Protective Measures for the Design of a Secure, Resilient and Cost-Effective Smart Grid System

by

Sohini Roy

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2022 by
Graduate Supervisory Committee:

Arunabha Sen, Chair
Anamitra Pal
Guoliang Xue
Martin Reisslein

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

Important features of smart grids are identified as efficient transmission of electricity and monitoring data, speedier recovery from disrupted power supplies, decreased operation and management costs, improved security, etc. All these can be made possible by a well-planned advanced communication system for the grid. However, most of the existing research not only fail to provide a clear understanding of the intra-and-inter dependencies of joint power-communication systems, necessary for a reliable and resilient operation of the grid, but also debates on the best suited design for the communication network. This dissertation introduces a simple, yet accurate multi-valued-logic based model of interdependency called the Modified Implicative Interdependency Model (MIIM) which can depict the interactions between the components of these power-communication systems and using this model an existing problem in the grid concerning cascading failure of entities is solved. Communication system for smart grid is responsible for securely sending both power transmission control data and environmental monitoring data to Control Centers. In this dissertation, a hybrid communication network, comprising of both wired and wireless communication is proposed together with a secure routing protocol to mitigate different types of cyber-attacks. Also, to prevent false data injections and owing to some limitations in MIIM, a further improvement is made to develop the Multi-State Implicative Interdependency Model which considers the data dependency of communication entities. In this dissertation, the issue of communication cost incurred due to ill-designed topology is also addressed, and an optimal-cost communication topology is

planned for modern smart grids. It is also identified that communication cost analysis cannot be done without considering the optimal Phasor Measurement Unit (PMU) placement problem. Consequently, the optimal PMU placement together with minimum cost network design problem is studied, and an attempt to minimize the overall cost is made in this dissertation. All the designs and network algorithms proposed here, are tested on substation location data of Arizona.

To Babai, Maa and Sattik

ACKNOWLEDGEMENTS

This is the time to look back at the bumpy way I have travelled in the last six years and the hurdles that at times seemed impossible for me to cross and take a sigh of relief as my last few steps will make me reach my goal. The first person, I would like to thank, with all humility, is myself; the timid, shy, introvert girl from a middle-class Indian family, who had shown the courage to leave behind her family, friends, culture, and country; cross the Atlantic and struggle alone in a new country about 8000 miles away from her home. It would not have been possible to complete this journey if that girl had lost her courage in the midway. The very next person I am indebted to is my husband, Sattik Chatterjee. Not only because he supported me and guided me in whatever step I took but also because, he is the only person who made me feel at home in this foreign country. I thank the love of my life who held my hands tight whenever I felt lost in this journey.

I am deeply thankful and grateful to my dissertation committee. Firstly, my Ph.D. advisor Dr. Arunabha Sen, without whose passionate guidance this dissertation would not have been possible. He made me step into this amazing world of network algorithms and has always encouraged me to think critically and work on new challenging problems. I fall short of words to thank Dr. Anamitra Pal, who not only enlightened me about power systems but also acted as my mentor throughout these years. Dr. Martin Reisslein is another person in my dissertation committee who inspired me quite often and taught me how a professor should always be available for the goodwill of his students. I would also like to thank my committee member, Dr. Guoliang Xue for all the valuable feedbacks on my

research work. I would also extend my gratitude to Dr. Geunyeong Byeon who helped me and worked with me on a part of the problem in chapter 7 of my thesis.

Then comes my lab mates and senior colleagues, especially Kaustav Basu, Suli Adeniye and Sandipan Choudhuri. The kind of cooperation, encouragement and affection I received from them is unforgettable. I thank my master's thesis advisor Dr. Uma Bhattacharya of IIEST who first motivated me to come to the U.S.A for pursuing my doctoral degree. I could have never dreamt this big without her.

I am highly obliged to my parents, Babai (Mr. Swapan Roy) and Maa (Mrs. Sipra Roy). They made me what I am today. I can never thank them enough. I also thank my sister Sayani, for listening to my research ideas every now and then and showing enthusiasm, even being from a completely different field of education, and also for maintaining the same bonding even being geographically apart. I also thank my in-laws, especially my father-in-law Shyamal Chatterjee for all his encouragements and efforts towards keeping me motivated throughout these six years. I must also thank my friends− Vaibhav, Garima, Moumita, Darpan, Dipanjan Da, Siddhant and Richa, who made my stay pleasurable in the U.S. Last but not the least, I would like to thank all my friends back in India, especially Bishakha, Tina, Poulomi, and Rumela. They are my support system and my soul-sisters. I thank them all for having faith in me and inspiring me throughout this journey. At the end, I feel humble to say that this dissertation is a fruit of the collective effort and good wishes of all these people mentioned above.

TABLE OF CONTENTS

Page

LIST OF TABLES…….. xi

LIST OF FIGURES……..... xii

CHAPTER

1. INTRODUCTION…….. 1
2. A SURVEY ON EXISTING INTERDEPENDENCY MODELS AND OPTIMAL COST COMMUNICATION NETWORK DESIGNS FOR SMART GRID…… 8

2.1. Interdependency Models……... 8

2.1.1. V. Rosato et.al Model…… 9

2.1.2. J. Wafler et. al Model……... 11

2.1.3. Xin. Liu et.al Model…… 15

2.1.4. Parvin Chopade et. al Model…… 17

2.1.5. Bamdad Falahati et. al Model……. 18

2.1.6. J. Sanchez et. al Model……. 20

2.1.7. IIM Model……. 22

2.1.8. Boolean Network Model…… 23

2.2. Importance of ICT Network Design for Smart Grid…… 26

2.2.1. Xingzheng Zhu et. al ICT Network Design…….. 28

2.2.2. Mostafa Beg Mohammadi et. al ICT Network Design…….. 32

2.2.3. Anamitra Pal et. al ICT Network Design……... 35

CHAPTER Page

2.2.4. F. Ye et. al ICT Network Design…… 39

3. MODIFIED IMPLICATIVE INTERDEPENDENCY MODEL…… 43

3.1. Designing of a Realistic Joint Power-Communication Network….. 45

3.1.1. Grouping Buses into Substations……... 45

3.1.2. Finding the Shortest Distance Between All pair of Substations and Selection of Control Centers……... 48

3.1.3. Placement of SADMs and Formation of SONET-Ring…… 49

3.1.4. Placement of OADMs and Formation of DWDM-Ring…… 50

3.2. Overview of MIIM and Modeling of IDRs…….. 51

3.3. Case Study…… 59

3.4. State Estimation Results……... 60

3.4.1. Overview of State Estimation…… 61

3.4.2. Hardware Failure of Gateway 13 and SADM 39 of IEEE 118-bus System……. 62

3.4.3. Damage of Substation 85 of IEEE 118-Bus System……. 64

4. IDENTIFICATION OF THE K-MOST VULNERABLE ENTITIES IN A SMART GRID…… 67

4.1. Problem Formulation…… 68

4.1.1. Inputs to the Problem……. 69

4.1.2. Decision Version of the Problem……... 69

4.1.3. Optimization Version of the Problem…… 69

CHAPTER Page

4.2. Integer Linear Program based Optimal Solution…….. 70

4.2.1. Variable List, Objective Function and Constraint Set…….. 70

4.3. Comparative Analysis between IIM, MIIM and Simulation Results…… 72

5. A SECURE SMART GRID MONITORING TECHNIQUE……. 79

5.1. Overview of the ICT Network Setup Phase for SSGMT……. 82

5.2. Risk Model and Assumptions for the SSGMT Network Setup…… 86

5.3. SSGMT Routing Scheme……. 88

5.3.1. Module 1: Data Forwarding to Substation Gateways by Sensors……... 88

5.3.2. Module 2: Data Forwarding by Substation Gateways to RSs and PDCs…… 88

5.3.3. Module 3: Data forwarding by RSs and PDCs to CC-gateways…… 91

5.4. Performance Evaluation and Simulation Results of SSGMT…….. 92

5.5. Overview of the Multi State Implicative Interdependency Model (MSIIM)……. 94

5.6. Routing of PMU Data from Substations to Control Centers Using MSIIM……... 98

5.6.1. Assumptions for Designing the Secure Routing Scheme using MSIIM……. 98

CHAPTER Page

5.6.2. Secure Routing Scheme using MSIIM…….

5.6.2.1. Module 1: Data Forwarding to Substation Gateways by PMUs…….. 99

5.6.2.2. Module 2: Data Forwarding by Substation Gateways to PDCs……... 100

5.6.2.3. Module 3: Data Forwarding by PDCs to CC-gateways…….. 105

5.7. Performance Analysis of MSIIM Simulation Results for the Routing Scheme……. 106

6. OPTIMAL COST NETWORK DESIGN FOR BOUNDED DELAY DATA TRANSFER FROM PMU TO CONTROL CENTER USING HIGH BANDWIDTH CHANNELS…….. 110

6.1. Problem Formulation…… 111

6.2. Difference between DCMT and RDCMST Problems…….. 113

6.3. RDCMST Integer Linear Program……... 117

6.4. Modified Prim (M_Prim) Algorithm…… 118

6.5. Evaluation of the Modified Prim (M_Prim) Algorithm……... 127

7. DELAY CONSTRAINED COMMUNICATION NETWORK DESIGN FOR PMU TO MULTIPLE CONTROL CENTER DATA TRANSFER.. 131

7.1. Problem Formulation…… 134

7.2. Optimal Solution for MRDCMSF Problem…….. 135

CHAPTER Page

7.3. Lagrangian Relaxation Based Solution…… 139

7.3.1. A Path Formulation for the MLCC…… 139

7.3.2. Lagrangian Dual…… 141

7.3.3. Subgradient Method for Solving the Lagrangian Dual…….. 142

7.4. Heuristic Solution for the MRDCMSF Problem…….. 144

7.5. Experimental Results…… 147

8. JOINT PMU PLACEMENT AND OPTIMAL COST NETWORK DESIGN FOR BOUNDED DELAY DATA TRANSFER FROM SUBSTATIONS TO CONTROL CENTERS…… 151

8.1. Joint PMU Placement and Communication Network Topology Design (JPMUPCNTD) Problem…….. 151

8.2. Optimal Solution for JPMUPCNTD Problem…….. 152

8.3. Experimental Results…… 158

9. CONCLUSION AND FUTURE DIRECTIONS…… 160

REFERENCES……... 164

LIST OF TABLES

Table		Page
3.1.	Truth Table for MIIM Operators……..	52
3.2.	Failure of Entities with Time obtained using MIIM……	59
3.3.	Failure of Entities with Time obtained using IIM……	60
5.1.	Parameter List for Simulation of SSGMT……	92
5.2.	Truth Table for MSIIM Operators……	96
5.3.	Evaluation of IDRs to Obtain State Values……..	96
5.4.	Parameter List for Simulation of the Routing Scheme using MSIIM……..	106
6.1.	Spanning Trees Corresponding to 4 Points in Fig. 6.1…….	114
6.2.	Ratio between the M_Prim and Optimal Solutions for Phoenix……..	128
6.3.	Ratio between the M_Prim and Optimal Solutions for Tucson……...	128
7.1.	H/Op And LUB/Op: Ratios Between The Heuristic And The Optimal; And Lagrangian Upper Bound And The Optimal, Respectively For The Tucson Data Set…….....	148
7.2.	H/Op And LUB/Op: Ratios Between The Heuristic And The Optimal; And Lagrangian Upper Bound And The Optimal, Respectively For The Phoenix Data Set……..	149
8.1.	Joint PMU Placement and ICT Network Design Cost for Different Smart Grid Systems……..	159

LIST OF FIGURES

Figure		Page
2.1.	State Machines for Components and Services and the Perception of their State in the Monitoring System……	13
2.2.	Meta Model for Smart Grid……	14
2.3.	Centralized Start Topology for Smart Srid……...	40
2.4.	Decentralized Start Topology for Smart Grid……..	40
2.5.	Centralized Mesh Topology for Smart Grid…….	41
2.6.	Decentralized Mesh Topology for Smart Grid…….	42
3.1.	Nomenclature for Every Entity of the Joint Network……..	44
3.2.	Substation Entities and Substation Division of IEEE 14-bus System….	46
3.3.	SONET-Ring Structure of IEEE 14-Bus System…….	47
3.4.	DWDM-Ring Structure of IEEE 14-Bus System…….	48
3.5.	State Estimation Result for Gateway 13 and SADM 39 Failure for Case1…….	63
3.6.	State Estimation Result for Gateway 13 and SADM 39 Failure for Case2…….	64
3.7.	State Estimation Result for Substation 85 Failure for Case1……...	65
3.8.	State Estimation Result for Substation 85 Failure for Case2……...	65
4.1.	Comparison Between Simulation, MIIM ILP and IIM ILP Results…….	73
5.1.	Critical Information Infrastructure Design for a Smart Grid of IEEE 14-Bus……..	83

Figure		Page
5.2.	Communication Delay vs. Malicious Nodes…….....	93
5.3.	Number of Compromised Nodes vs. Packet Drop……	94
5.4.	Flowchart Describing Module 2 of Secure Routing Scheme……...	103
5.5.	Identification of FDI Attack by PDC…….....	104
5.6.	Percentage of Node Compromise vs. Communication delay……...	106
5.7.	Percentage of node compromise vs. Percentage of Packets Dropped…..	107
5.8.	Number of fabricated packets vs. Average energy consumed……..	109
6.1.	RDCMST Problem Instance with 4 Points in a 2-Dimensional Plane……	114
6.2.	Opt_RDCMST Solution with 11 Points……...	123
6.3.	M_Prim_RDCMST Solution with 11 Points……...	124
6.4.	Blocks and Diagonals in the RDCMST Problem Instance……..	124
6.5.	Distance between Points on the Upper and Lower Diagonals…….	125
6.6.	Distance between Points in Arizona…….	125
6.7.	M_Prim_RDCMST on Phoenix $DS_{2,2}$ with $\delta = 50$…….....	128
6.8.	Opt_RDCMST on Phoenix $DS_{2,2}$ with $\delta = 50$…….....	128
6.9.	Opt_RDCMST or M_Prim RDCMST on Phoenix $DS_{2,2}$ with $\delta \geq 58.35$……	130
7.1.	SS to CC Direct Connections……...	132
7.2.	SS to CC through LCC…….	132
7.3.	A Possible Solution for the MRDCMSF Problem….....	133

Figure Page

7.4. Substations (red points) in the Intersection (contention) Area of Several q-circles, where Each q-circle Correspond to an LCC……. 144

8.1. Network connecting Substations and Central Control Center(s)…….. 152

8.2. Power Layer Graph for IEEE 14-Bus Smart Grid…… 154

8.3. ICT Layer Graph for IEEE 14-Bus Smart Grid…… 158

Chapter 1

INTRODUCTION

Maintaining a sustainable lifestyle is contingent upon an uninterrupted supply of electricity. Modern power utilities try to ensure the continuity of this supply by running an intricate network that consists of intra-and-inter-dependent power and communication system entities. For example, the power network measurements of the smart grid obtained by its sensors must be transferred to the control center by the communication entities. At the same time, the communication network entities themselves need power from the smart grid for their continued functionality. This interdependency has become critical in a smart grid environment where the failure of an entity in one network can lead to failures of the entities of the other network. Thus, it is essential to understand the interdependencies between the two types of networks for predicting the effect of failure of one or more entities on the overall system state. An inaccurate prediction may impact the decision making of an operator which can then lead to a less efficient operation of the grid. Also, the design of the communication network of the smart grid should be robust, scalable, and reliable enough to support the smooth functioning of the power grid, for example, correct predictions from the operator should reach the substations and get executed properly. The communication from the opposite direction, that is from the Remote Terminal Units (RTUs) placed at every substation or Phasor Measurement Units (PMUs) placed at particular substations, should also be fast, secure and cost-effective.

Models that have been proposed previously to describe the intra-and-inter dependencies of critical infrastructures, such as [1], [2] often lack physical realism as they are too simple to correctly portray the complex structure of the interdependent networks [3]. A specific drawback pertaining to the electrical infrastructure is the lack of clarity in the description of its communication network design. For example, in [4] a design of the joint network was given for the IEEE 14-bus system. However, the details of the Information and Communication Technology (ICT) network were missing. The Implicative Interdependency Model (IIM) [5] was successful in representing the complex interdependencies of a joint network using simple yet accurate Boolean logic-based Inter-Dependency Relations (IDRs). Yet, the main drawback of the IIM model is that it can only represent the operational/non-operational state of an entity by assigning 0/1 value to each entity in the IDR. However, it cannot capture the inherent nature of reduced operability of the smart grid entities. As a result, resolving IIM IDRs often resulted in inaccurate calculation of state values of entities. IIM also failed to accurately model the communication network entities as it lacked knowledge of the communication network design. A few such interdependency models are studied and discussed in the Chapter 2 of this dissertation.

This dissertation attempted to address all the above-mentioned requirements. In Chapter 3 of this dissertation, a two-layered Multi-Valued-Logic based dependency model named as the Modified Implicative Interdependency Model (MIIM) and its application is presented. With the help of a power utility in the U.S. Southwest, in Chapter 3 of this dissertation, a realistic design of the structure and operation of the power-and-

communication network of a typical smart grid is presented. In Chapter 1, a modified version of the IIM, termed as the Modified Implicative Interdependency Model (MIIM) [6] is proposed which uses this realistic communication network design coupled with standard IEEE bus systems for demonstrating a smart grid. A summary of the differences between IIM and the proposed modified IIM (MIIM) using the concept of Inter-Dependency Relations (IDRs) and performance of the two models during a system failure is also shown in the chapter.

In Chapter 4 of this dissertation, identification of the K-most vulnerable entities in a smart grid system is done using the MIIM proposed in Chapter 3. Since the entities in both power and communication network exhibit complex intra-and-interdependencies between them, the failure of one or more entities can lead to subsequent failure of multiple other entities leading to a catastrophe. In order to avoid such a condition, the researchers should have a clear understanding of such complex dependencies between the entities and based on that they should be able to identify the most critical entities in the smart grid system, failure of which can maximize the network damage. Efficient hardening techniques [7] followed for such critical entities can save the smart grid from a huge damage. Yet, in order to identify the most vulnerable entities in the system, clear understanding of the design of the joint power-communication system as well as an appropriate interdependency model to capture the complex dependencies in a smart grid are necessary. The MIIM model proposed in Chapter 3, serves this purpose. It considers different operational levels of the entities and models the complex dependencies using Multi-Valued Logic based equations called Interdependency Relations (IDRs) where the entities in the joint network are

considered as logical variables. Therefore, just by solving these IDRs the Smart Grid Operator (SGO) can identify the operational states of different entities in the network after some initial failure has taken place in the system and thereby recognize the most vulnerable entities in the network. Now, even after identifying all the vulnerable entities in the system, the SGO can have a budget constraint of hardening only *K* entities of the network, where K can be any integer. In that case, it is important to identify the *K*-most critical entities in the system. The problem of identifying the *K*-most vulnerable entities in a joint power-communication network is already proved to be NP complete in [5]. Therefore, an Integer Linear Programming (ILP) based solution for the problem is given in Chapter 4 using the MIIM IDRs.

Bearing in mind, both the advantages and disadvantages of both wired and wireless communication networks, in Chapter 5 of this dissertation, a hybrid communication network design comprising of both wired and wireless communication technologies is proposed and a secure routing protocol for such networks is also given. An improved communication network transforms a traditional power grid system into a smart grid by incorporating features like full-duplex communication between the ICT entities, automated metering in the smart homes, power distribution automation, and above all intelligent decision making by means of pervasive monitoring of the power system and thereby securing its stability. Therefore, it is beyond any question that the ICT of a smart grid must be accurate, scalable, and secure enough to instantly identify any kind of abnormal behavior in the entities of the power network; securely communicate that information to the CC and thereby help in taking necessary and timely action to ensure uninterrupted

power supply. Although, the best suited design for the communication network of a smart grid is still not very clear and also monitoring of the smart grid environment is another important aspect which is often overlooked while designing the ICT system for the smart grid. Therefore, in Chapter 5 a hybrid ICT network design is presented that is responsible for both environmental monitoring as well as health monitoring [8] of the smart grid entities using advanced sensor-based network and also wired communication system, coupled with a secure routing scheme for that ICT network. Also, in Chapter 3, by solving the IDRs it cannot be predicted if the data received from an operational entity is correct or it has false data injected into it. CCs completely depend on the data carried to it by the communication system from the PMUs to make all the required analysis. However, such data dependency is not covered in the dependency model MIIM. On the other hand, cyber-attacks like False Data Injection (FDI) are very common in the communication system of a smart grid. Thus, in Chapter 5 the MIIM is further modified to provide a model named Multi-State Implicative Interdependency Model (MSIIM) which together with the structural and functional dependencies also considers the data dependency between the ICT entities of the smart grid. A novel multi-path data routing technique is also proposed in this chapter which not only identifies False Data Injection (FDI) attacks but also detects the source of the attack using MSIIM IDRs.

Communication network topology design problem in a Smart Grid environment, where electric power transmission grid control data, generated by the Phasor Measurement Units (PMUs), needs to the exchanged between substations (SS) and Controls Centers (CC) in real time, has received considerable attention in the research community in recent times

[9]−[11]. In [9] the authors analyze topologies to provide technical guidance to power utilities, for the design of the communication network for the PMU based real-time applications. Many researchers have studied the optimal PMU placement problem with specific objectives, such as, full network observability. However, optimal PMU placement without taking into account the cost of ICT infrastructure, may lead to the design of a very expensive Wide Area Monitoring (WAM) system, where the cost of ICT infrastructure may dominate the cost of the PMUs. In [10], the authors studied joint optimization of PMU placement and associated CI cost. This line of research is continued in [11], where the authors study the optimal PMU communication link placement (OPLP) problem that simultaneously considers the placement of PMUs and communication links for full observability. In Chapter 6, the network design problem studied in [9] is formalized as the Rooted Delay Constrained Minimum Spanning Tree (RDCMST) problem and unlike the existing works on this area, it is studied in a geometric setting, instead of a topological setting. Owing to the problem being NP complete, not only an ILP based solution is proposed in this chapter but also a heuristic solution is given for the RDCMST problem. The heuristic solution is named as Modified-Prim's Algorithm (M_Prim), which is a modified version of the well-known Prim's algorithm for construction of a Minimum Spanning Tree of a graph.

In Chapter 7 of this dissertation, the results presented in Chapter 6 are further extended by considering Multiple Local Control Centers (MLCCs) where data from every PMU in the SSs must arrive at one of the multiple LCCs within the specified delay threshold. This setting gives rise to a new problem, where a Delay Constrained Spanning

Forest needs to be created instead of a Delay Constrained Spanning Tree. The notion is formalized with the introduction of the Multi-Rooted Delay Constrained Minimum Spanning Forest (MRDCMSF) problem. In Chapter 7, an optimal solution for the problem is provided using ILP, a Lagrangian Relaxation based solution and also a heuristic solution with an innovative contention resolution mechanism.

Now, just as communication network topology design for the smart grid environment has received attention from the researchers in recent years, similarly there has been an extensive number of studies on optimal PMU placement problem [12]. Yet, most of the studies in optimal PMU placement and ICT network topology was done in isolation. Only in the recent years, researchers have pointed out that conducting these two studies in isolation may not really lead to total ICT infrastructure design cost minimization. In order to minimize total infrastructure design cost, in Chapter 8 of this dissertation both PMU placement and ICT topology design problems are considered simultaneously. An ILP based solution for the problem is proposed in Chapter 8 and results are given considering real substation locations of Arizona.

Finally, in Chapter 9, the dissertation is concluded, and future work directions are discussed.

Chapter 2

A SURVEY ON EXISTING INTERDEPENDENCY MODELS AND OPTIMAL COST COMMUNICATION NETWORK DESIGNS FOR SMART GRID

2.1. Interdependency Models

In recent years, critical infrastructure operators, government agencies responsible for ensuring the proper functioning of these infrastructures, and infrastructure researchers have been highly recognizing the fact that national critical infrastructures such as power grids, communications networks, transportation systems, and water distribution systems are strongly dependent on each other. It is not only true that these systems cannot perform in isolation but also, they are closely connected with each other and improvement in one such infrastructure can have huge effects on the other one. Now, each of these individual infrastructures are themselves highly complex. Therefore, simultaneously exploring interdependency between all the infrastructures is a herculean task. Hence, for the purpose of this study, the focus is on two of the most critical physical infrastructures – *power* and *communication*.

As we concentrate exclusively on the power and communication networks, it is observed that the power grid entities like buses, transmission lines, transformers etc. depend completely on the communication network comprising of ICT entities like servers, gateways, sensors, communication channels, routers etc. for their own health monitoring [8] and also for receiving control messages or operational commands. Similarly, the ICT system rely on the power grid for power supply. Thus, a failure in one or more entities of

a particular network can trigger failures in the other network and eventually lead to a catastrophe. Therefore, the complex interdependencies between the two types of infrastructures should be identified properly to monitor their operations.

For the past few years, different researchers have taken different approaches to understand the interdependencies between the critical infrastructures. Some have adopted a detailed way of exploring the dependencies yet, others have abstracted the interdependent systems in some way to simplify the representation of the complex dependencies between them. Although, in the abstract approach, a simplified representation of the interdependencies was achieved, finding the appropriate level of abstraction became another challenge for the researchers. In this chapter, a brief survey on the existing interdependency models is presented and the drawbacks pertaining them is identified.

2.1.1. V. Rosato et.al Model

The V. Rosato et. al [1] uses a coupling model to analyze the interdependencies between power and communication networks. This model analyzed the power flow and the SCADA data flow separately. The authors in [1] have considered the high voltage Italian electric transmission network (HVIET) as the power network and the high-bandwidth backbone of the Italian Internet network (GARR) as the communication network to study the power-communication interdependencies.

In this model, the HVIET power grid is represented as an undirected graph consisting of *N* nodes and *E* arcs. Each node is categorized into one of the three types, namely− *S nodes* or Source nodes which inserts power into the grid, *L nodes* or Load nodes

that draw out power from the grid and *J nodes* or Junction nodes which belong to neither of the two other categories but transmits power from one node to the other. The transmission lines are represented as edges in the graph. A DC power flow model [13] that provides a linear relationship between the active power flowing through the lines and the power drawn by nodes, is used here to determine the electrical flow in the grid. After every failure in the system, the power flows are recalculated using [13]. A ratio of the change in total power drawn by the *L* nodes after a failure to that drawn before the failure determines the Quality of Service for the power grid in this model.

In order to model the GARR communication network, a graph consisting of the Italian universities and research institutes as nodes and the Italian high-bandwidth backbone of the internet dedicated to linking them as edges is used. A probabilistic packet routing model is used here for sending a generated packet to a randomly selected node taken as the destination. Average delivery time is defined as the average packet forwarding time from source to destination across all packets delivered correctly within a given time interval. This average delivery time is then used as a metric to define the efficiency of the network for a particular value of λ, $(0 \leq \lambda \leq 1)$ which is the probability that a node will generate a packet and it is taken into account at each time step to calculate the total amount of traffic that flows into the network.

This coupling model now represents the dependency of ICT entities on the power grid entities in the following manner. Any communication entity or a node in the graph used to represent the GARR is assumed to draw power from its closest *L* node '*i*' in the power graph. If the power drawn by an ICT node before any failure in the system is denoted

by $P_i^{(0)}$ and the power drawn by the same ICT entity after the failure and after revaluation of the power flows as P_i then, the ICT entity will remain operational or in an "on state" if $P_i \geq \alpha P_i^{(0)}$, where α is indicated as the strength of coupling between two networks.

This coupling model is then used to simulate and analyze the effect of arbitrary link failures in the power grid with the value of α taken as 0.75. Their most important discovery obtained from their simulation, states that even with a small failure within the HVIET grid, like having even a negligible ratio between the number of tripped transmission lines to the total number of transmission lines, the ICT network getting totally disconnected, is not only unrealistic but also gives a wrong evaluation of the interdependencies between the two types of networks. The approach of isolated modeling of the two interdependent networks done in [1] gives a very limited idea about the actual behavior of smart grid systems. Not only it violates the basic concept that interdependent networks cannot be studied in isolation but also while trying to establish the coupling it only takes into account the unidirectional relationships from power to communication networks, while fully ignoring the effect of failures of ICT entities on the power grid.

2.1.2. J. Wafler et. al Model

The dependency model presented in [14] considers both the ICT network components and the power grid components to study the dependencies in smart grid and also depicts the role of ICT networks on smart grid. The authors in [14] have divided smart grid entities into five different categories namely−Category A containing power entities which do not need or have any communication means or any software assistance, for

example power lines and mechanical devices; Category B contains power entities that can be configured but they run without any aid, for example distribution energy resources; Category C contains power components which are totally software controlled and has full dependence on ICT entities like the Intelligent Electronic Devices (IEDs) like PMUs; Category D consists of all the software-enabled communication devices like routers and finally Category E has such ICT entities which are not dependent on software, for example communication cables. It is also pointed out that overlap of categories in some of the smart grid entities is inevitable and those entities are mainly responsible for cascading failures. According to this model, smart grid services mainly run on the basis of B, C and D components as the other two categories only contain hardware devices.

After categorization of the entities, the authors define state machines for these categories and also for the smart grid services. It is mentioned that separate states for the examined failure modes must be created for a quantitative analysis, and transition rates or probabilities must be ascribed to the transitions. Two operational states are defined for hardware entities, and they are '*ok*' or operational and '*F*' or failed. Operational states of Category B, C and D entities can be categorized into− '*ok*' ensuring full functionality of the entities; 'active error' or E_a state, in which errors made by entities are persistent but may or may not result in the failure of the component and until then it has no direct effect on the operational states of other components; 'passive error' or E_p state in which errors are short term and the entity often goes back to 'ok' state in this case; 'active failure' or F_a state in which an entity remains active and has direct contributions on behaving as a failed entity, for example sending wrong monitoring data or wrong control command, and it is

beyond any question that this state of a component can trigger E_a, E_p or F_a states in other entities dependent on it; the final state described in this model is 'passive failure' or F_p state which denotes a state where a component is not active or not-responsive when it is needed and this state can also trigger F state of hardware entities and also F_p state in other entities.

The novelty of this model lies in the fact that, together with smart grid entities, the operational states of smart grid services is also considered here. Fig. 2.1. shows the state machines for all the different components of smart grid considered in this model.

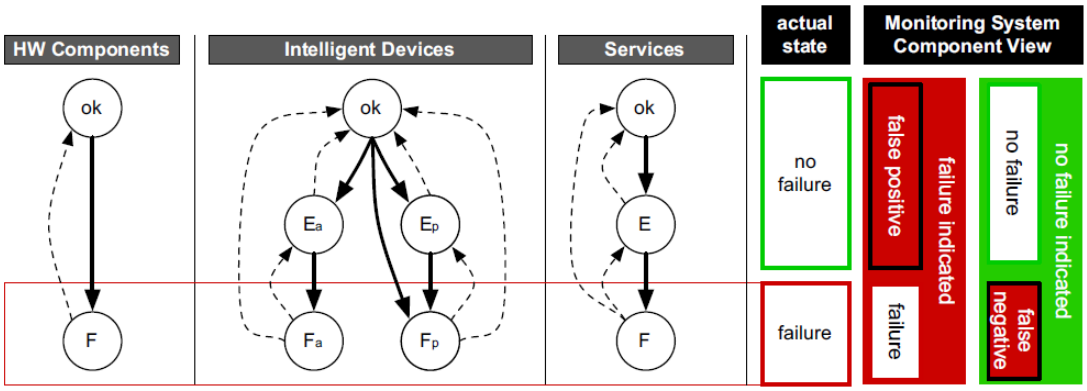

Fig. 2.1. State Machines for Components and Services and the Perception of their State in the Monitoring System [14]

Influence of each type of component mentioned in fig. 2.1. on the other is also detailed by the authors in this model. The main idea presented in [14] while studying the influence of each entity on other is that errors cannot have a direct influence on the operational states of other entities, yet failures will definitely have. The total monitoring system is perceived as a service in this model which can decide if a failure has taken place or not, but it can also have errors or failures, and this may result in false positives and false negatives.

A couple of techniques for quantitative analysis of the model is given in [14], although no such analysis is presented by the authors. Yet, a meta-model is proposed to describe the state of the whole system for the CC. This meta-model is a collection and interpretation of data from the monitoring system used to evaluate the system's criticality level. The states of power grid as a whole and ICT as a whole is considered in two axes here, as shown in fig. 2.2. The state of services is coupled with the power grid entities state to determine the overall state of the power grid. The states of ICT components and services, on the other hand, are utilized in conjunction with a logic that determines which services are vital to define the state of the ICT system.

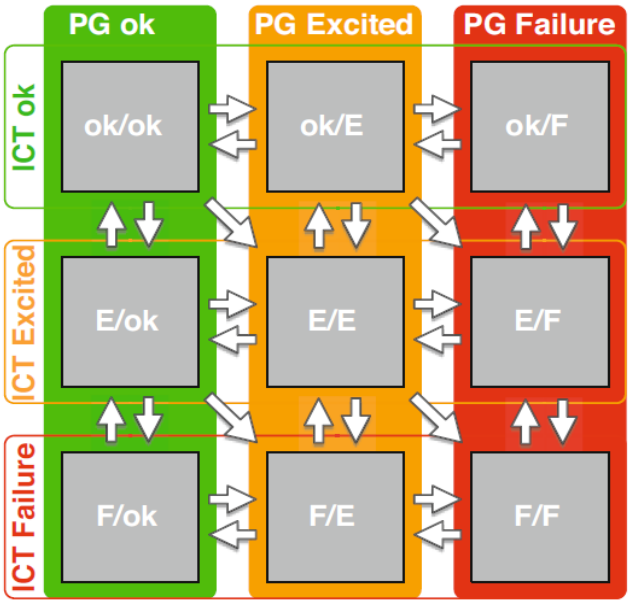

Fig. 2.2. Meta Model for Smart Grid [14]

A service-based approach is followed in the model where 'Failure' means a service is not being delivered and immediate action is required, 'Excited' stands for the state where the service is on the verge of failure and '*ok*' indicates full functionality. When the PG is excited that indicates that even though all customers are powered, the system is in an

excited state like N-1 redundancy is harmed or load is critical. PG failure indicates one or more customers is disconnected from the power supply. ICT excited on the other hand states that even though all critical ICT services are delivered correctly, the system is on the about to fail and it may be the case that non-critical entities already failed or there is congestion in the system. ICT failures can indicate unresponsiveness or incorrect operations of ICT entities. Intersection of the two axes in the meta-model gives rise to nine operational states of the system as shown in fig. 2.2.

The good side of this model is that it gives a nice categorization of the smart grid components and the states at which each of these components can be. Most other models do not take services as a separate component and often ignore the errors which are considered as a separate state in this model. Yet, analysis of this model is not presented by the authors and a very high-level view of the system is considered. It is also not clear how the dependencies can be represented and stored in the system for analyzing the overall system state as mentioned in the meta model.

2.1.3. Xin. Liu et.al Model

In the literature [15], a two-layer distribution system model with power-communication network entities is proposed. This model is then used to formulate a restoration process for a routing problem considered for the smart grid communication network. The authors in [15] have pointed out the need for having a fine-grained model of the power-communication network by taking into account actual components of the two

networks. The MIIM model proposed in Chapter 3 of this dissertation also identified the same requirement and therefore a granular consideration of the smart grid entities is done.

[15] also proposes a distribution system recovery model by considering a realistic distribution grid model, which is converted from the referenced models like IEEE-123 or Ckt7 systems. The entities which are considered in this model includes−power nodes, branches, switches and substations. Distribution Automation (DA) which is mainly enabled with remote systems like automated switches is essential for an intelligent distribution system. A wireless overlay is created by the authors based on a mesh network for the distribution grid model presented in [16]. In [15] network equipment controls DA services and that equipment depend on the power grid for power supply. A Mixed Integer Linear Programming (MILP) based optimization problem is formulated here for maximizing the restored energy within a limited window.

The IEEE-123 bus system is considered for the power grid design and then an algorithm to synthetically construct the communication network for a given grid is presented here. Even though detailed description as well as illustration of both the power and communication layer of this model is given in [15], the authors do not elaborate on how the intendencies between the two layers is captured. The routing model given here gives a restoration process for the smart grid and the aim here is to minimize the energy loss over time. Yet, the order of restoration for communication entity failure is ignored here. The restoration process discussed in this literature is out of scope for this dissertation, but it is evident from the literature that even in very recent research like [15] (published in

May 2022) the interdependency model presented is not able to fully capture the interactions between the entities in the two layers.

2.1.4. Parvin Chopade et. al Model

In [17] a quantitative vulnerability assessment of critical infrastructure systems is provided. The authors have highlighted the fact that SCADA systems should be coupled with proper security mechanisms in order to provide reliability and efficiency to the grid. Major cyber threats pertaining the SCADA system, or the ICT network of a smart grid is pointed out in [17]. One of the major drawbacks of most of the existing ICT network designs for smart grid is that the security of the network is not taken into account when such systems suffer from a number of vulnerabilities. In order to express the interdependency between the power and the communication network, just like MIIM [6] and IIM [5], in [17] also a two layered graph is considered. As described in [17], the vulnerability of a critical system is related to the likelihood that a disruption will result in a consequence Q which can either be societal or technical, but that should be greater than a critical value q during a given time period T. If the consequence of a disturbance that occurs at time t is $Q(t)$, where $t \in T$, then the vulnerability of the critical infrastructure system is given by the probability $P(Max_{t\in T}Q(t) > q)$. This vulnerability analysis can aid in the formulation of solutions to potential crisis circumstances and provide a foundation for prioritizing different options for improving system performance.

The authors in [17] also calculated values of the topological characteristics of the networks and compared their error and attack tolerances, or how well they perform when

vertices are deleted randomly or maliciously. A smart grid of IEEE 14-Bus system is considered for illustration of the topological analysis and factors like *average path length* indicating number of edges in the shortest path between two points, *clustering coefficient* denoting the ratio between present number of edges to possible number of edges between two points, and *degree distribution* representing the number of edges connected to a vertex; are taken into account to characterize the structure of interdependent networks. In this model not only details of the ICT network are missing but also it is not clear how after a vulnerability analysis situations are handled. In this graph-based model, nodes with higher degree distribution are considered more critical but that may not be the case in reality.

2.1.5. Bamdad Falahati et. al Model

In [2], the authors have highlighted a concept called indirect dependencies between power and ICT networks. According to the authors, direct interdependency describes the circumstance in which a cyber network failure causes a power network element to operate wrongly or cease to function entirely. On the other hand, indirect interdependencies have a different and more complicated impact on power system reliability than direct interdependencies. In case of indirect dependencies, failure of a set of elements in one network do not cause the failure of or modify the behavior of the elements in the other network directly and instantly, but that failure will have an impact on the performance of the elements in the other network when further failures occur in the system. Based on this definition, both hidden and unacknowledged failure can be categorized into indirect dependency, which can lead to failures in the monitoring system or the protection system

of the grid. In [2], a reliability assessment algorithm is proposed to model the indirect dependencies between power and communication network.

Indirect dependencies are further categorized as−(i) Indirect element-element interdependencies (IEEI) which takes place when ICT devices that are physically/logically connected to a power system element fail; and (ii) Indirect network-element interdependencies (INEI) which emerge when a failure does not occur on any cyber devices that are directly connected to the power network, but instead occurs within the cyber network, affecting the power element's performance when failures occur in the future. In [2], an indirect link that represents an indirect interdependency between a cyber entity γ and a power entity δ is shown as $\Gamma = (\gamma : \delta)$. Availability of a power network element is represented here in terms of failure rate and repair rate of the power device. It is mentioned that the presence of an indirect interdependency between a cyber element and a power element leads to a degradation of the availability of the power element if the cyber element is not available.

The authors have used the concept and formulations of state updating based on probability of states obtained on the basis of failures in the cyber network. Running two interconnected and heterogeneous networks becomes conceivable with this state update method. The influence of indirect cyber-power interdependencies on reliability indicators such as the loss of load probability (LOLP) and expected energy not served (EENS) is quantified using an algorithm. However, just like most other interdependency models, this model also considers only two operational states of the elements. Just like IIM [5], this model also considers 0 state for fully operational entities and 1 state for non-operational

entities. This dissertation identifies this consideration as a major flaw in the smart grid interdependency models.

2.1.6. J. Sanchez et. al Model

In [4], in order to represent coupled infrastructure interdependencies, the authors provide a topologic-driven technique that assesses complex-weighted networks to measure some topological indices like Betweenness Centrality and Efficiency. The method is demonstrated in [4] using a typical French distribution network as well as an ICT network in the surrounding area.

The primary issue addressed in [4] is a lack of tools for analyzing and studying linked critical infrastructures, specifically to detect their interdependencies and vulnerabilities in the context of widespread ICT deployment. The answer to this challenge may result in a clearer picture of the system-of-systems, which will aid in dependability, security, and risk assessments, as well as making power systems more secure. After careful analysis of different existing methods, the authors decided to go with Complex Network which is a graph-based tool for modeling complicated systems that has been used to examine and understand massive systems with complex topologies and hidden interdependencies. This method can be used to determine the topology and connectivity attributes of a system, as well as to analyze failure and cascade occurrences. Critical Infrastructures can be depicted as a Complex Network of interconnected systems that operate together to accomplish a common goal.

In the methodology discussed in [4], a network is defined as a set of interconnected items and a graph is considered as a mathematical representation of it. Complex Networks theory is the application of graph theory to the study of vast and complex systems. Vertices indicate system elements like buses, routers, servers etc. in the smart grid and edges denote relationships or connections between the vertices. It is to be noted that in this case the authors are considering both physical and logical connections between vertices as edges. The whole smart grid system is represented as a graph $G = (V, E)$ in [4] where V is set of all vertices and E is the set of all edges. In order to evaluate the interdependencies between the different types of vertices, the edges in the graph are considered as representations of interdependencies here and they are classified into four types as follows−Type 1: a directed connection between two electrical nodes, for example: power transmission lines; Type 2: a directed connection between two ICT nodes, for example data transmission lines; Type 3: A directed edge from an electrical node to an ICT node, for example: energy supply line to ICT entities; and Type 4: a directed connection from an ICT node to an electrical node, for example an edge used to send commands to the electrical components. Now an adjacency matrix for this graph is created where each element gets a value as follows:

$$a_{hj} = \begin{cases} 1, & if\ the\ link\ (h,j)\ is\ type\ 1\ or\ 3 \\ 1i, & if\ the\ link\ (h,j)\ is\ type\ 2\ or\ 4 \\ 0, & otherwise \end{cases}$$

According to the authors in [4], this adjacency matrix helps in analyzing the main complex network indices like node-degree, betweenness centrality and efficiency. Even though, it is true that considering the interdependent networks as complex networks and thereby representing them as directed graphs can help in understanding the complex network

indices, this form of representation of the smart grid network or any other coupled network is an overly simplistic approach. The type 1 and type 3 links or type 2 and type 4 links cannot be considered as same in reality. The type of interactions between the nodes is very essential and should be taken into account while modeling the dependencies.

2.1.7. IIM Model

The Implicative Interdependency Model (IIM) [5] was successful in representing the complex interdependencies of a joint network using simple yet accurate Boolean logic-based Inter-Dependency Relations (IDRs). In IIM [5], the smart grid system can be viewed as a multilayer network, represented as a set $J\big(E, F(E)\big)$, where $E = P \cup C \cup CP$ which means all entities belonging to power layer, communication layer and the intermediate layer joining power and communication entities; and *F(E)* represents the set of IDRs. The entities in power layer (layer 1) are considered as P type entities where $P = \{P_1, P_2, \dots P_n\}$ and entities in ICT layer (layer 2) are named as C type entities where $C = \{C_1, C_2, \dots\ C_m\}$. The set *F(E)* is used in both the models to capture the dependencies among interacting entities in the network. It is to be noted here that only structural dependencies are considered to generate the IDRs in IIM. This model has a binary nature and the entities in that model can either be operational with a state value of 0 or be non-operational with a state value of 1. The most common feature of reduced operability in critical infrastructures is ignored in IIM.

Now, in order to describe the model, if C_i, an entity of layer 2, is considered which will be operational if (i) C_j which is another entity of layer 2 and P_a which is an entity of

layer 1, are operational, or (ii) C_k which is an entity of layer 2 and P_b which is an entity of layer 1 are operational, and (iii) C_l which is an entity in layer 2 is operational. Then the corresponding IIM IDR for C_i would be: $C_i \leftarrow \Big(\big(C_j \,.\, P_a\big) + \big(C_k \,.\, P_b\big)\Big) \,.\, C_l$. In this IDR, '.' denotes logical AND operation and '+' denotes logical OR operation. Similarly, the IDR for a P type entity can also be expressed.

Initial failure of entities in one or both the network layers in the IIM model will cause cascading failure of more entities in both layer until a steady state is reached. Induced failure occurs when an entity fails after the initiation of the failure process is already done by other entities. In IIM, failure occurs in unit time steps, with the first failure occurring at time step $t = 0$. The effect of entities killed in prior time steps is captured in each time step. Major drawbacks of the IIM model are that it is overly simplistic and is not able to capture any operational state that is in between purely operational or failed state. This model has also completely ignored the complexities of the Information & Communication Technology (ICT) and never discussed how actual failures of ICT entities can affect the power grid entities.

2.1.8. Boolean Network Model

In a Boolean network model [18], a network is represented as a graph $G = (V, F)$ where the nodes V represents elements of the network and F defines a topology of edges between the nodes and a set of Boolean functions. The set of nodes can be represented as $V = \{v_1, v_2, \ldots v_n\}$, and each node v_i has a function associated with it which takes the states of all nodes connected to v_i as input. The state of a node v_i at time t is denoted as: $x_i(t)$.

The value of the states can either be 0 denoting false or 1 denoting true. Each such node v_i is associated with a Boolean Function BF_{v_i}. That Boolean Function can be logical operations like AND, OR, NOT etc. The state of a node $v_i : x_i(t+1)$ at time (t+1) can be determined by executing such logical operations that the BF_{v_i} comprises of. For example, $x_i(t+1) = BF_{v_i}(x_j(t), x_k(t), x_l(t))$; or, $x_i(t+1) = x_j(t)ORx_k(t)AND(NOTx_l(t))$, where $x_j(t), x_k(t)$ and $x_l(t)$ are the states at time t, of the nodes v_j, v_k and v_l connected to the node v_i.

The differences between Boolean network [18] model and the Implicative Interdependency Model (IIM) [5] can be given as follows:

i. In IIM, only logical AND and OR operators can be used to form the Inter-Dependency Relations (IDRs) between the entities. Yet, in Boolean Network logical NOT is also allowed.
ii. In IIM, 1 indicates a non-operational state and 0 indicates operational state. However, it is just the opposite in a Boolean Network. Therefore, the goal of the operator handling the two network models should be totally opposite. The operator handling an IIM based network will try to minimize the system states at any point of time by adopting entity hardening methods. On the other hand, the operator using a Boolean Network will try to maximize the system state so that lesser number of entities are in the false state 0.
iii. In IIM, an entity which has failed at time t can never be operational at time $t' > t$. With time, only more and more entities can fail, until and unless that entity is repaired or replaced physically. In Boolean network, a node whose state value

has turned false or 0 at a time t may regain a state value 1 after the failure of certain other entities at time $t' > t$. For example, at time t, the node v_i having a state value defined by the following equation: $x_i(t+1) = x_j(t) OR x_k(t) AND (NOT x_l(t))$ fails since the state value of node v_l has become 1. Now, again at time $t' > t$, the node v_l fails again and its state becomes 0. Then at time $t' > t$, the state value of v_i will change back to 1.

iv. In IIM, due to absence of the NOT operation, the cascade of failure can propagate in one direction only. For example: If we have a dependency as follows: $A \longleftarrow B, B \longleftarrow C, C \longleftarrow D$ then, if A fails, it will lead to the failure of B and that will lead to the failure of C and then D will fail. Therefore, the cascade propagates in the direction: $A \rightarrow B \rightarrow C \rightarrow D$. Even if there are circular dependencies like: $A \longleftarrow B, B \longleftarrow C, C \longleftarrow D, D \longleftarrow A$, then after the failure of A, the IDRs will be updated as: $A \longleftarrow B, B \longleftarrow C, C \longleftarrow D$ and the propagation of cascade will be unidirectional only. However, in a Boolean Network, there can be dependencies like: $A \longleftarrow NOT(B), B \longleftarrow NOT(A)$. Here the Boolean functions are shown as dependencies. Now, after the failure of entity A, other Boolean functions cannot be updated and they will still have the state of A as a variable, the value of which can change the state of other entities. In this case as A fails, B will become operational and as B fails, A will become operational. Therefore, the flow of cascade will keep moving back and forth as the system state of A and B changes.

v. If we consider the Boolean Functions for A, B, C and D nodes as the following dependencies: $A \leftarrow NOT(B), B \leftarrow (C) AND\ (D), C \longleftarrow D, D \longleftarrow A$, then, if A is

false, D will be false, C will be false, B will also be false. Just as B becomes false, A becomes true. Now, as A is true, D is true, C is also true, B also becomes true. Just as B becomes true, A changes from true to false. Therefore, there is a continuous toggling of states between the nodes. This situation can never arise in IIM.

2.2. Importance of ICT Network Design for Smart Grid

With the continuous technological advancements taking place, communication plays a huge role in making the power systems more reliable. It is to be noted that only an improved Information and Communication Technology (ICT) can transform a traditional power system into a smart grid system. Smart grids are obtained by incorporating features like full-duplex communication between the ICT entities, automated metering in the smart homes, power distribution automation, and above all intelligent decision making by means of pervasive monitoring of the power system and securing its stability. Therefore, it is beyond any question that the ICT network of a smart grid must be accurate, scalable, and secure enough to instantly identify any kind of abnormal behavior in the entities of the power network, securely communicate that information to the control center (CC) and thereby help in taking necessary and timely action to ensure uninterrupted power supply.

The Supervisory Control and Data Acquisition (SCADA) have been in practice for a long time. In order to improve the remote monitoring of the power network and fructify the concept of a smart grid, current and voltage sensors are further updated to develop more advanced devices like Phasor Measurement Units (PMUs) which have a sensing module,

a data processing module, a memory and a communication module. Output from the processing module of the PMU is sent to the communication module and finally to the network to be sent to the control centers (CCs). These PMUs sample current and voltage phases at the rate of 48 samples per cycle and send 30 samples per second. CCs use these PMU data to perform a number of analytical tasks like state estimation, to estimate voltage stability margin, to validate generator model, generate a contingency list for the network and so on.

The failure in communication entities can have a massive impact on power system operations. For instance, during the 2003 blackout in United States, a failure in alarm software led to the human operators being unaware of the transmission-line outage. The transmission-line outage eventually led to a cascading failure of power systems eventually causing the blackout.

Health monitoring of the power grid is done by means of state estimation [6] which is one of the key components of the energy management system (EMS). Loss of measurements might affect the performance of state estimator. A badly estimated state can result in wrong decision making by the operator. Loss of measurements from the sensors in a bus, can result in bad estimate of the state of that bus or the neighboring buses. In such a case, the estimate in that bus or the neighboring buses will be less reliable in comparison to the other buses. This brings the need for a fast and secure communication network to carry the PMU data from the substations to the CCs with minimum delay, minimum cost and in a secure manner. Although, designing of a robust ICT network for smart grid has

become a boiling topic of research, the best suited ICT design for a smart grid is still not very clear.

In this chapter some existing optimal cost communication network designs for the smart grid system is studied and it is observed that most of the works focus entirely either on optimal PMU placement, or on connecting the substations with minimum cost, or on reducing the delay. However, none of them tries to address everything simultaneously.

2.2.1. Xingzheng Zhu et. al ICT Network Design

[10] states that the high cost of PMUs requires an optimal placement of PMUs which can ensure full observability of the system. PMUs are responsible for collection of phasor data (system states) from the power entities, but those data should be sent to the CC for analyzing. A suitable communication network should be there to support this. This literature [10] aims at finding the optimal PMU-communication link placement which takes into account both optimal PMU placement and communication link (CL) placement, such that the power system is fully observable. In order to achieve this goal, the communication capability needed for each such communication link is also analyzed here. The model proposed in this paper reduces the installation cost to a great extent as compared to the traditional optimal PMU placement methods.

The problem addressed in [10] is stated as follows. Given a power network, the optimal PMU placement for the network and optimal communication link for the PMUs placed should be determined; such that the total power network is observable, but the installation cost of the communication network is minimized. In the optimization version

of the problem, the minimum set of *K* PMUs in the power network *N* (defining the number of buses in the power grid) needs to be computed, such that data from each bus is measured and the installation cost (per length cost + bandwidth cost) of the minimum set of *L* communication links that connect the PMUs among themselves and also with the PDC, is minimized as well.

A power network comprising of buses, transmission lines and transformers; type of each bus (generator bus, load bus or zero-injection-bus); a connectivity matrix *M* defining the connection between the buses which takes a value 1 for each $m_{i,j}$ if bus *i* and bus *j* are connected by a transmission line/transformer or if bus *i* and bus *j* are same*;* and takes a value 0 otherwise; are taken as inputs to the problem.

The solution to the problem outputs− (i) PMU installation vector: $\mu = [\mu_1, \mu_2 \ldots \mu_n]$ where μ_i gets a value 1 is a PMU is installed on bus *i* and 0 otherwise, (ii) location of communication links (CL); i.e. a network topology (in the form of a matrix) showing how the PMUs are connected among themselves and with the PDC, and (iii) bandwidth requirement of each such CL.

The main objectives realized in [10] includes optimal PMU placement, optimal communication link placement on the basis of communication capacity or bandwidth requirement of the communication links, minimizing the installation cost which comprises of both the cost per length of the communication channel as well the cost per Gigabytes (GB) of data transferred and also obtaining full observability of the power system. In order to achieve these objectives, the authors in [10] made a number of assumptions for the joint power-communication system. It is assumed that there is only one Phasor Data

Concentrator (PDC) placed in the Wide Area Monitoring System (WAMS). This may not be a correct assumption for larger WAMS, but for the sake of understanding their approach this can be considered. It is also assumed that the requirements on the CL capacity depend on the PMU data traffic and the data routing scheme. The different choices on transmission paths for PMU data can affect CL cost. The bandwidth cost (cost/usage ─ which is the usage cost/GB and also cost/Bandwidth─ which denotes the connection speed that is bandwidth cost per megabyte and so on). In [10], mainly cost/GB is considered, that constitutes a significant proportion of the total system installation cost. The shortest CL does not necessarily indicate minimum cost. This means, even if the shortest distance CL is selected but the speed required for the CL is very high then the cost/GB will be high, and the overall installation cost will increase. It is stated that the presence of a greater number of Zero-Injection-Buses (ZIB) with no generation and no load, will reduce the number of optimal PMUs to be placed in the network as there is no need to collect data from a ZIB. As a result, the cost on CLs and the amount of data traffic is also reduced. It is considered that there may exist some CLs in the power system before the CL placement is designed by this model. With the given model, new CLs may be installed into the system as well as more capacities to some of the existing CLs can be provided. For the transmission of all PMU data, a CL may belong to multiple data transmission paths. It is assumed in this work that the time division multiple access (TDMA) scheme is adopted to avoid collision of data transmission from numerous PMUs over a common CL.

In order for the ILP proposed in [10] to give the required outputs, a number of constraints need to be satisfied. For the power system to be fully observable, the

observability of bus *i* denoted by g_i should be greater than *k*, where *k* is the number of PMUs responsible for monitoring each bus in the system. If *N* denotes the power network. g_i denotes the observability of bus *i*. It is calculated as: $g_i = \sum_{j \in N} m_{i,j}\, \mu_j + \sum_{j \in N} m_{i,j}\, z_j\, y_{i,j} \quad \forall i \in N; \sum_{i \in N} m_{i,j}\, y_{i,j} = z_j \quad \forall j \in N$. Here, the term $y_{i,j}$ denotes credit that bus *j* can put on bus *i* and its value is 1 if bus *j* is a ZIB and 0, otherwise. Here, z_j is 1 if the bus *j* has total of 1 credit that can be assigned to itself or its adjacent buses and z_j is 0, otherwise. Now the observability of bus *i* is calculated as the summation of the number of PMUs placed on each bus *j* of the network, summed with the credit that bus *j* can put on bus *i*. Now, each PMU has a capacity of the number of buses it can monitor. This capacity is incorporated in the observability calculation in the following way, $g_i = \sum_{j \in N} m_{i,j} \omega_{i,j}\, \mu_j + \sum_{j \in N} m_{i,j}\, z_j\, y_{i,j} \quad \forall i \in N$. Here $\omega_{i,j} = 1$, if bus I can be measured by the PMU located at bus j and 0 otherwise. The designed bandwidth for each CL should be greater than or equal to the bandwidth requirement for that CL. The total number of buses a PMU at bus *j* can measure, must be lesser than a given maximum ω_j^{max} which denotes the maximum capacity of that PMU. A CL should be placed for the branch *i-j* if the bandwidth required for that CL is greater than *d*. Here *d* denotes the minimum unit of data traffic that a PMU can generate. A CL will only be updated with new capacity if the current capacity is less than the bandwidth requirement for that channel. There should be at least one available path to efficiently communicate the data from the PMU at bus *i* to the PDC. In order to fulfil the transmission requirements of all PMUs in the system, the system bandwidth should be sum of all PMU requirements.

2.2.2. Mostafa Beg Mohammadi et. al ICT Network Design

The approach proposed in [11], aims at optimizing the cost at different portions of the WAMS. In order to do that, the cost of optimal PMU placement, placement of a PDC and also the communication infrastructure associated with them are considered simultaneously and minimized. Binary imperialistic competition algorithm is used for the optimal placement of PMUs. Dijkstra' s single source shortest path algorithm is used to find the minimum cost for the communication links, and it is also used for the optimal placement of the PDC. The method proposed in this paper considers the factor that there can be buses with PMUs already placed on them and some portion of the power network can already have some communication links. Unlike [10], [11] considers only distance to find the minimum length of communication link and thereby select that link to optimize the cost. However, the cost of communication link may not only depend on the length. The bandwidth cost is also important. A shortest length communication link may not always guarantee the minimum bandwidth requirement for that link and therefore the overall cost can be high. This factor is not considered in [11]. However, in [10], the optimal placement of PDC is not considered. This factor is considered in this work. Like [10], [11] also considers the presence of some preinstalled communication links but together with that it takes into account of the fact that there can be some preinstalled PMUs as well.

The problem addressed in [11] states that given a power network, the optimal PMU placement for the network, optimal communication links to be placed and optimal placement of PDC are to be determined, such that the full power network is observable with respect to state estimation and the overall installation cost of the WAMS is minimized.

A power network comprising of buses, transmission lines and transformer; type of each bus (generator bus, load bus or zero-injection-bus); a connectivity matrix A defining the connection between the buses just like [10], are taken as inputs to the problem.

The problem outputs– (i) a PMU installation vector: $X = [x_1, x_2 \dots x_n]$ where x_i is defined as

$$x_i = \begin{cases} 1 & if\ a\ PMU\ is\ installed\ on\ bus\ i \\ 0 & Otherwise \end{cases},$$

(ii) location of the PDC, (iii) location of communication links (CL); i.e., a network topology, in the form of a matrix, showing how the PMUs are connected among themselves and with the PDC.

The main objectives of [11] includes optimal PMU placement, optimal communication link placement, optimal placement of PDC and minimizing the overall installation cost of the WAMS and full observability of the power system. It is ensured that there is only one Phasor Data Concentrator (PDC) placed in the Wide Area Monitoring System (WAMS). The communication system is designed in [11] on the basis of the OSI model layers. Cost of the communication infrastructure is simplified in this work as: $Cost_{total} = Cost_{active} + Cost_{passive}$. This denotes the summation of the cost of active and passive components. The active components are the devices such as network routers, switches etc. The passive component is length of the optical fiber channel. Reducing the cost of both the components can reduce overall cost. The imperialistic competition algorithm proposed here is used to decide which of the buses should be placed with a PMU and which other buses will be monitored by PMUs placed in adjacent buses. After the

optimal placement of PMUs is done, the location of the PDC is determined by placing the PDC at each bus at a time and calculating the shortest distance of the PDC from the PMUs using Dijkstra's algorithm. After the distances are calculates, the cost is estimated for placing the communication links. This process is repeated by placing the PDC is all the buses. Finally, the location of the PDC is determined on the basis of minimum cost as determined by calculating the shortest path from all the PMUs. It is to be noted here that this process is time consuming and difficult to apply on larger systems.

In order to produce the required solution, the following constraints, mentioned by the authors in [11], need to be satisfied. The phasor voltage of each and every bus is accessible at least in one way by the optimal placement of PMUs, in other words full observability of the network should be ensured. Each PMU should be connected to the PDC via a fiber optic cable. A PMU should not be placed on a bus with only one incident line (transmission line), i.e., a PMU should not be placed on a bus which is connected to only one other bus in the network. (If the power network is considered as a graph, then the pendant vertices should be avoided for placing a PMU). The reason is that if a PMU is installed in such a bus, it can only measure the voltage phasor of that bus and the one which is incident to that bus. In other words, by using a PMU it is possible only to measure two voltage phasors. It is also stated that a PMU should not be installed in a zero-injection-bus, as the zero-injection bus will not have any data to send to the control center. The WAMS should withstand (N-1) contingency like a failure of a PMU or a transmission line etc. Therefore, each bus should be either monitored by at least two PMUs. In case, bus is

observed by only one PMU then by some other means the bus should be observable like using Kirchhoff's law from adjacent zero-injection buses etc.

2.2.3. Anamitra Pal et. al ICT Network Design

The authors in [12] have pointed out the fact that the cost of upgrading a substation is much larger than the cost of an individual device and this has emerged as the primary constituent of the total expenditure for PMU placements. Therefore, for an optimal PMU placement scheme not only the number of PMUs that need to be placed should be taken into account but also it should be considered that how many substations are being disrupted for the upgradation purpose. [12] presents an Integer Linear Programming based methodology for PMU placement scheme while considering realistic costs and practical constraints. Factors like successful integration of renewable energy sources with the high voltage network, minimum substation disruption and availability of dual-use line relays (DULRs) acting as PMUs or availability of branch PMUs etc. are also considered here.

While the [10] and [11] focus mainly on the placement of minimum number of PMUs to get full observability and also on selecting the minimum cost communication links to connect the PMUs among themselves and with the PDC, [12] brings another factor for estimating the cost incurred with PMU placement and that is the cost of disrupting a substation. The lesser number of substations are disrupted for the PMU placement operation, the better is the operation of the power grid. If a greater number of substations are shut down for the PMU placement operation, then the power supply will be highly affected.

The problem statement of [12] states that, given a power network with substation division for the buses, the optimal branch PMU placement for the network should be done while minimizing the total cost which is calculated as the sum total of the cost of the PMU devices and the cost of disrupting a substation from its normal function due to the PMU placement.

Inputs to the problem includes a graph $G(V, E)$, where each vertex $v_i \in V$ represents a bus in the power grid and each edge $e_{ij} = \{v_i, v_j\}$ represents a transmission line or transformer connecting the buses i and j. The node set V is partitioned into $k \geq 2$ blocks: $\{B_1, B_2 \dots B_k\}$ where each block represents a substation. There are no edges in between the vertices of each block. Thus, nodes in each block form an independent set and the input to problem is a k-partite graph G. The problem outputs the number of branch PMU or DULRs placed and the location of the DULRs or vertices having a DULR placed.

The main objectives of [12] includes optimal number of branch PMU placement or DULR placement and minimization of the total cost of PMU placement which is equal to the sum total of the device cost and the cost incurred in disrupting a substation from its normal operation due to the PMU placement operation. The objective function is given as: $Minimize\left(\sum_{i=1}^{k} c_i y_i + \Delta \sum_{e \in E}\{w_e^h + w_e^l\}\right)$ where c_i is the cost for disrupting a block and Δ is the cost of a DULR. y_i is 1 if a block or substation i is disrupted and 0 otherwise. w_e^h is 1 if a DULR is placed at the high end of the edge e and 0 otherwise. w_e^l is 1 if a DULR is placed at the low end of the edge e; 0 otherwise.

It is assumed here, that each vertex v_i is denoted by two integers: $v_i = \{x_i, y_i\}$ where x_i denotes the block number or the substation number which the vertex v belongs to and y_i denotes the index number within each block. However, it is not mentioned in [12] whether the block number or the index number decides the ID of the vertex. A vertex is denoted as v_i but what does the *i* refer to is not clear. Considering a block B_2 with 5 buses in it and the block number as 2, each bus within that block will have the same x_i value which is equal to 2 but they will have different y_i value ranging from 1 to 5. In that case what will be the *i* of that v_i is not mentioned. It is not specified whether the value of *i* be just the bus number, irrespective of the x_i and y_i values. In that case, different variables should be used as index of *v*, *x* and *y*. Neighborhood (N_v) of node *v* is defined by the number of vertices adjacent to the vertex *v*. The authors stated that a DULR must be placed on either end of an edge. A DULR can observe both the vertices v_i and v_j when placed on an edge $e_{ij} = \{v_i, v_j\}$. If a DULR is placed at the v_i end of the edge, then the substation containing v_i will be disrupted and same is true for the other end as well.

[12] mentions of a number of constraints that need to be satisfied in order to obtain the solution for the problem. The constraints are given as follows. It is mentioned that different buses have different priorities with respect to system stability or security. Those buses with higher priority should be given higher preference while placement of PMUs. A substation can have one or more buses in it. The main goal of this scheme is not to minimize the number of buses on which PMUs should be placed but to minimize the number of substations which will be disrupted for one or more PMU placements. A DULR can measure voltage of both ends of a transmission line. Thus, DULRs should be used for

ensuring observability and redundancy. Also, a DULR must be placed at one of the ends of an edge. Each vertex *v* should be observed by at least one DULR placed on one of the edges that incident on the vertex. This is represented by the constraint $\sum_{e\in E_v}\{w_e^h + w_e^l\} \geq 1$ where E_v denotes all the edges incident on vertex *v*. If a DULR is placed at the v_i end of the edge $e_{ij} = \{v_i, v_j\}$, then the substation containing v_i must be disrupted and same is true for the other end as well. Together with all these basic constraints, some additional constrains should also be satisfied— (i) Redundancy should be provided to the critical buses. The constraint $\sum_{e\in E_v}\{w_e^h + w_e^l\} \geq t + 1 \ \forall v \in C$, ensures this for all vertices *v* which are critical vertices ($v \in C$), and there should be a total of $(t + 1)$ DULRs monitoring the vertex *v* where *t* is the maximum number of DULRs that may fail. Therefore, each critical vertex *v* should be observable under $(N - t)$ contingency. A practical constraint is there which states that certain substations cannot be disrupted from normal operation for a PMU placement. Each such substation *i* will have $y_i = 0$ in the ILP formulation, where y_i is denoted by 1 if a block or substation *i* is disrupted and 0 otherwise. Some of the substations may have preinstalled PMUs. If $P \subseteq V$ is a subset of vertices where PMU is already placed then, each such vertex $v_i \in P$ may observe M_{v_i} other vertices where $M_i \subseteq N_i$ and N_i is the number of neighbors of $v_i \in P$. Then, the placement of new DULRs should be concerned about the observability of only $V - M_P$ number of nodes or in other words the nodes that are currently not observable. M_P is denoted as: $M_P = \bigcup_{1\leq i\leq |P|} M_i$ which denotes the total number of vertices that are already observed. Any edge that joins two nodes inside a block should not be considered for PMU placement. However, if this elimination of edges makes a node disconnected from the rest of the network, then that

node is placed inside a dummy block and the edge of that node with other nodes in the real block are considered and a DULR must monitor the vertex in the dummy block. In this case, the cost of PMU placement is calculated as the cost of the DULR summed with the cost of disrupting the real block. The authors also state that the presence of Zero-injection buses should be considered. Yet, the iso-voltage zero-injection buses (ZIB connecting two buses with same voltage) should be treated differently from the ZIBs connecting two buses with different voltages. At least one incident edge of one of the buses connected by iso-voltage zero-injection buses should be placed with a DULR. However, how the zero-injection buses connected to another bus with different voltage is handled is not very clear. Since the paper mentions about only PZIs or iso-voltage zero-injection buses which is a subset of the TZI (total number of zero-injection buses in the system) and representing the zero-injection buses connecting buses with same voltages; it can be assumed that the other zero-injection buses are not treated as zero-injection buses or in other words they are treated as normal injection buses, and a DULR placement is not ignored for such buses.

2.2.4. F. Ye et. al ICT Network Design

[9] uses Network Simulator 3 (NS3) to simulate different possible communication topologies for the smart grid system with PMUs being the sender nodes and the CC being the receiver. The authors have pointed out the importance of finalizing the topology of the ICT network that exchanges real-time data between the CC(s) and the SSs. A new algorithm is also proposed in [9] to optimize the clustering of substations in such a way that data latency performance requirements of decentralized communication topology can be achieved.

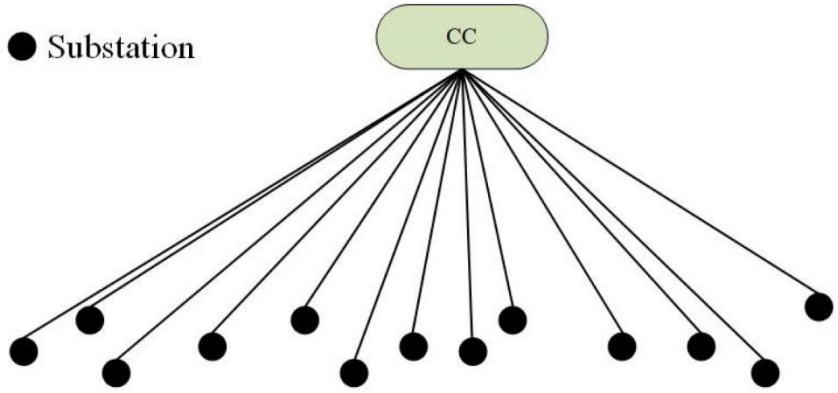

Fig. 2.3. Centralized Start Topology for Smart Grid [9]

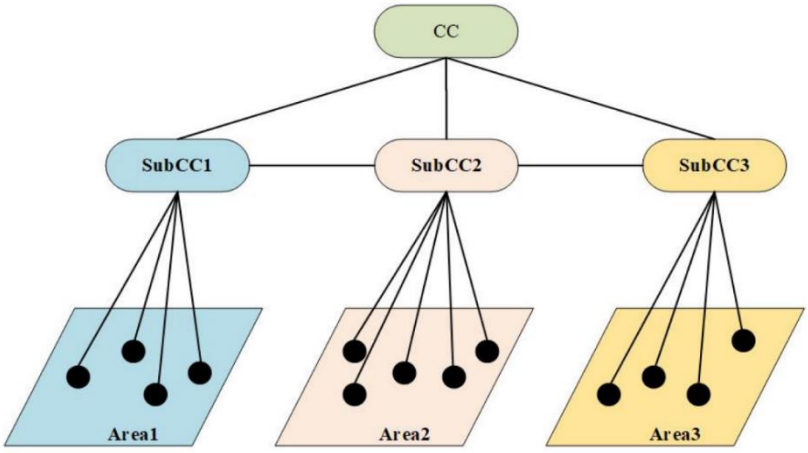

Fig. 2.4. Decentralized start topology for smart grid [9]

Two main types of communication network topologies are studied in [9] namely−star network and mesh network. In a start communication network, all the communication nodes are directly connected directly to the target node. In a power system scenario this means all SSs are connected directly to the CC. Fig. 2.3. shows this structure.

It is mentioned by the authors that, in order to construct such a network topology, new communication lines and towers are necessary, and it is already implemented in some real power systems. The fig. 2.3. shows only a centralized version of this topology where

all the SSs are directly sending their data to the CC. However, a decentralized design as in fig. 2.4. would be more realistic to portray a real smart grid star topology. In that decentralized design, data from SSs first go to one of the sub-CCs acting as hubs and those sub-CCs forward data to the main CC. On the other hand, a mesh network, defines a network in which the communication structure is similar to the mesh structure of power transmission lines. A mesh topology for ICT systems in smart grid can also be centralized as in fig. 2.5. and decentralized as in fig. 2.6.

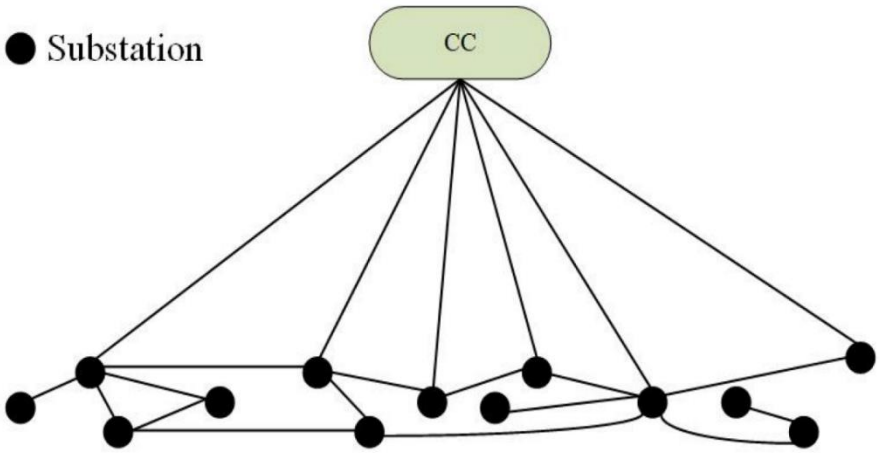

Fig. 2.5. Centralized Mesh Topology for Smart Grid [9]

The advantages and disadvantages of all these four types of topological design is discussed in [9] and finally it is concluded by the authors that the decentralized communication network topology can be designed at a lower cost by optimizing the substation cluster boundaries. This can be a good choice and help PMU based applications achieve good performance by limiting data latency to an acceptable level. The simulation results in [9] also show that better performance can be achieved by mesh network at a lower cost since it has built-in redundancy and does not always need new communication links

and towers. Therefore, mesh networks are considered by the authors as a better choice for ICT designers who focus more on reliable communication and has a limited budget. Although, this statement is self-contradictory as mesh network already has so much in-built redundancy that a budget constrained optimal communication network design cannot be done if mesh topology is considered. According to the results in [9], the performance of a star network is the highest, but it also requires adding redundancy in order to improve fault-tolerance. The authors state that this network topology is more suitable for those ICT designers who desire the highest quality of communication performance and do not have any budget constraints. Yet, star networks can be considered when a specific budget is there and then redundancy can be added while satisfying the budget constraints. When the design budget is very low, star topology will ensure at least a single connection between a SS and the main CC but if mesh topology is still considered then all substations may not be connected to the main CC. Depending on the available budget, the ICT designers can gradually move from a star topology to a full mesh or even a semi-mesh topology.

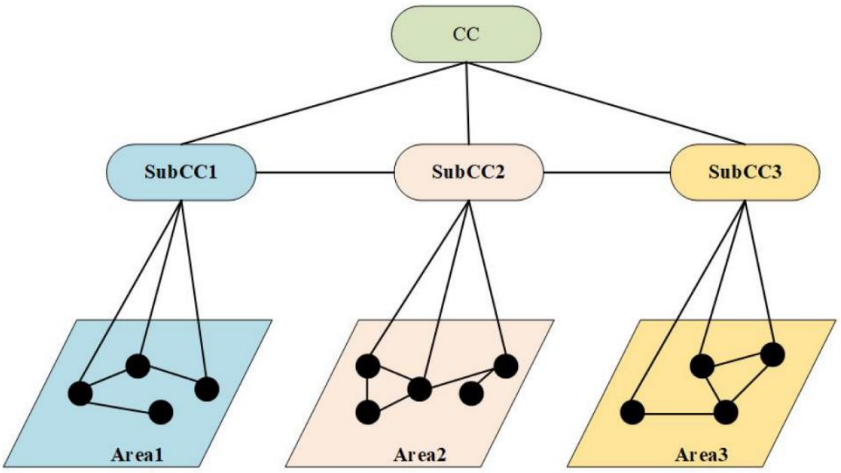

Fig. 2.6. Decentralized Mesh Topology for Smart Grid [9]

Chapter 3

MODIFIED IMPLICATIVE INTERDEPENDENCY MODEL

The Implicative Interdependency Model (IIM) [5] discussed in Chapter 2 was successful in representing the complex interdependencies of a joint network using simple yet accurate Boolean logic-based Inter-Dependency Relations (IDRs). However, it also failed to accurately model the communication network entities as it lacked knowledge of the communication network design. With the help of a power utility in the U.S. Southwest, this chapter presents a realistic design of the structure and operation of the power-and-communication network of a typical smart grid.

In this chapter, the smart grid is viewed as a multilayer network, where entities in power layer (layer 1) are called P type entities, $P = \{P_1, P_2, \dots P_m\}$, entities in communication layer (layer 2) are called C type entities, $C = \{C_1, C_2, \dots\ C_n\}$, and entities which belong to both the layers (layer 3) are called CP type entities, $CP = \{CP_1, CP_2, \dots CP_o\}$. Fig. 3.1. classifies the joint network entities into these three categories. The figure also provides subdivisions of each of the three types of entities and the nomenclature assigned to them.

In fig. 3.1, the P type entities are subdivided into buses, transmission lines/transformers, and battery backup. The C type entities are subclassified as substation entities (Type 1), synchronous optical networking (SONET)-ring entities (Type 2), or dense wavelength division multiplexing (DWDM)-ring entities (Type 3).

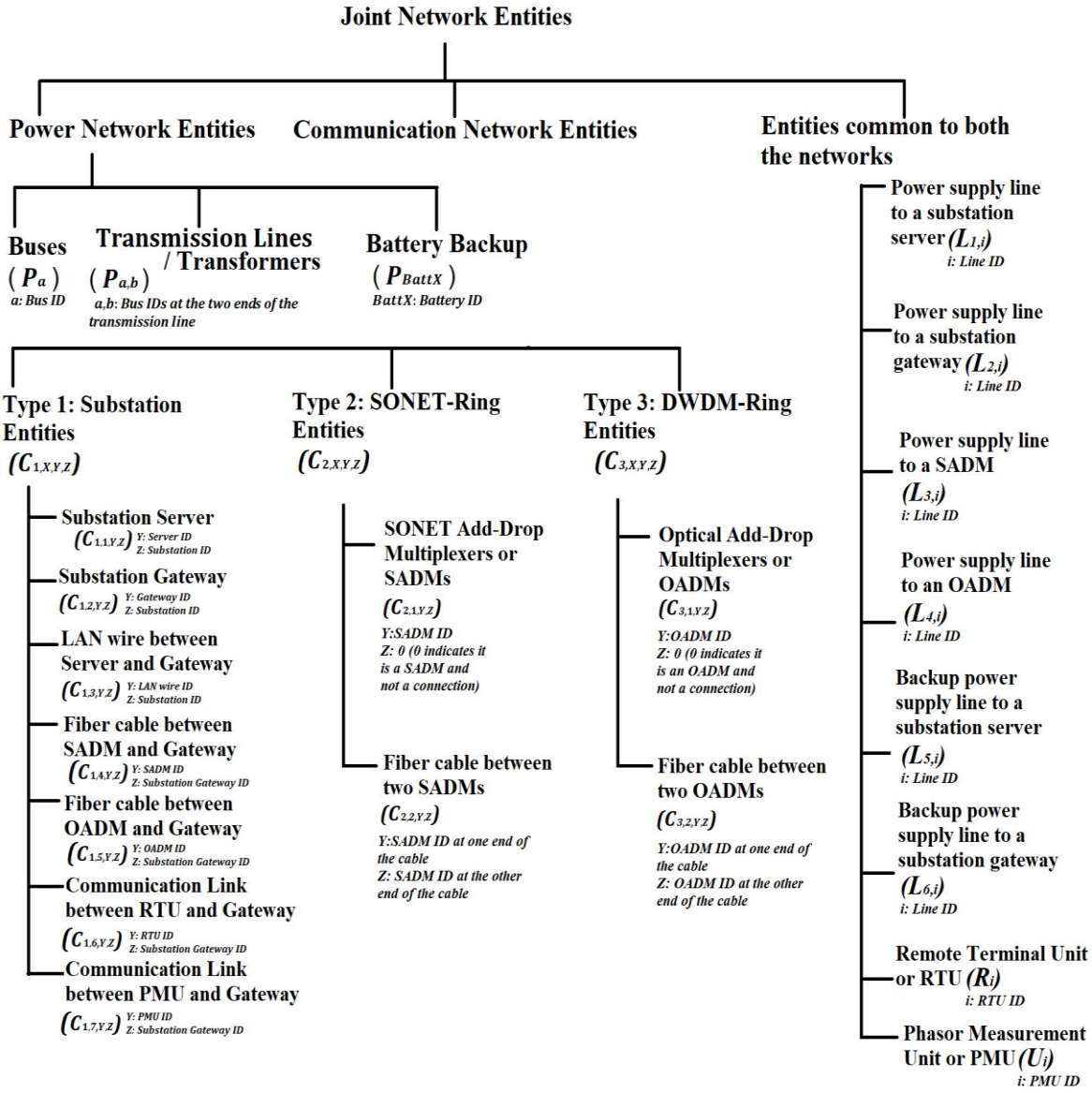

Fig. 3.1. Nomenclature for Every Entity of the Joint Network

Subdivisions of each of the three types of C type entities are also shown in fig. 3.1. The CP type entities consist of L type entities (power supply channels to different C type entities), R type entities (corresponding to remote terminal units (RTUs)), or U type entities (corresponding to phasor measurement units (PMUs)). The design principles of the joint network are explained in section 3.1. in more details.

3.1. Designing of a Realistic Joint Power-Communication Network

3.1.1. Grouping Buses into Substations

The buses of the power network are grouped into substations based on the logic given in [19]. The substation specific communication entities are Type 1 entities of fig. 3.1. This step is further subdivided into the following sub-steps:

The first sub-step is placing substation servers and gateways. The substation server $(C_{1,1,Y,Z})$ is the main computing device of a substation. The supervisory control and data acquisition (SCADA) system inputs from RTUs and the synchrophasor system inputs from PMUs, reach the substation server via the gateway $(C_{1,2,Y,Z})$. The substation servers of the control centers use SCADA/PMU data to perform state estimation. Substation servers of other substations compress and encrypt SCADA/PMU data to forward them to the substation gateway. The gateway then sends the SCADA and PMU data to the control centers through the low bandwidth optical channels using SONET over Ethernet (SONETоE) [20] and high bandwidth optical channels using Ethernet over DWDM (EoDWDM) [21], respectively. Hence, the gateway connects the substation server to the rest of the communication network outside the substation and also to the PMUs and RTUs within the substation. Any data coming to and going from the substation server must pass through the gateway. The gateway also has a firewall that protects the server from cyber-attacks. The server is connected to the gateway via LAN connection $(C_{1,3,Y,Z})$.

The second sub-step is supplying power to the Type 1 ICT entities. The substation server and gateway receive power from the buses inside the substation. In order to avoid

power outage within the substation, a battery backup (P_{BattX}) is also present in every substation. The battery supplies power to the Type 1 ICT entities when the buses in the substation do not have power.

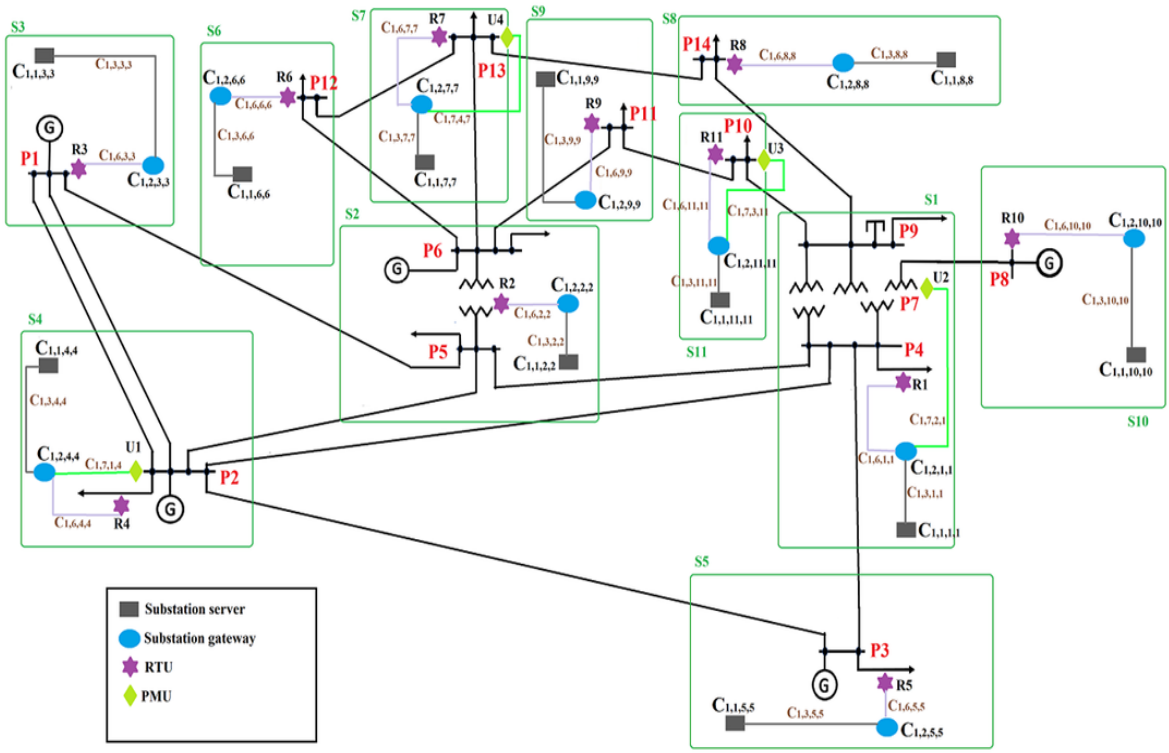

Fig. 3.2. Substation Entities and Substation Division of IEEE 14-bus system

The third sub-step is placing two geographically diverse fiber optic cables from each substation. There are two types of fiber optic channels going out from the gateway of each substation. One is the low bandwidth cable $(C_{1,4,Y,Z})$ that uses SONEToE technology, and the other is the high bandwidth cable $(C_{1,5,Y,Z})$ which uses EoDWDM technology. In order to observe the performance of the synthetic network under different scenarios, two different cases are considered in this chapter with respect to data transfer via the optical fiber cables. In Case 1, the SONEToE channels are responsible for carrying RTU data to

the nearest SONET-add-drop-multiplexer (SADM) of the SONET-ring (elaborated in section 3.1.3.) while the high bandwidth EoDWDM channel can only carry PMU data to the nearest optical-add-drop-multiplexer (OADM) of the DWDM-ring (elaborated in section 3.1.4.).

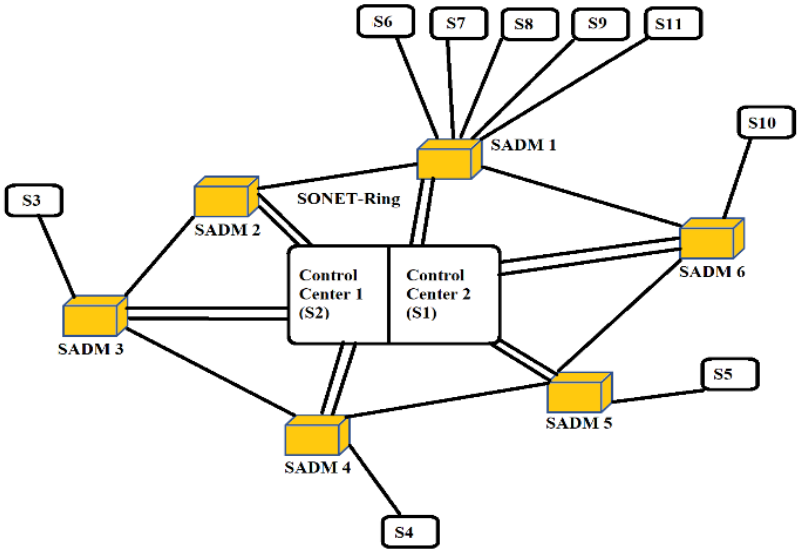

Fig. 3.3. SONET-Ring Structure of IEEE 14-Bus System

In Case 2, under normal conditions, the low bandwidth SONEToE cable is responsible for carrying the RTU data to the nearest SADM of the SONET-ring while the high bandwidth channel is responsible for carrying the faster PMU data to the OADM of the DWDM-ring. However, in case of failure of the low bandwidth channel, in Case 2 (unlike Case 1), the EoDWDM channel can transmit SCADA inputs from the gateways to the SADMs. For fault tolerance, the control center gateways are connected to every SADM in the SONET-ring via multiple low bandwidth channels and also to every OADM in the DWDM-ring via multiple high bandwidth channels. As an illustration, fig. 3.2. shows the substation division of the IEEE 14-bus system along with the substation servers and

gateways. The $(C_{1,4,Y,Z})$ and $(C_{1,5,Y,Z})$ cables are shown in fig. 3.3. and fig. 3.4., respectively.

The final sub-step is placing RTUs and PMUs. Every substation has RTUs (R_i). However, due to budget constraints, PMUs (U_i) are placed in only some of the substations using the methodology proposed in [12]. R_i and U_i measure SCADA system input data and synchrophasor system input data from the buses inside the substation and send them to the substation gateway via communication channels $(C_{1,6,Y,Z})$ and $(C_{1,7,Y,Z})$, respectively.

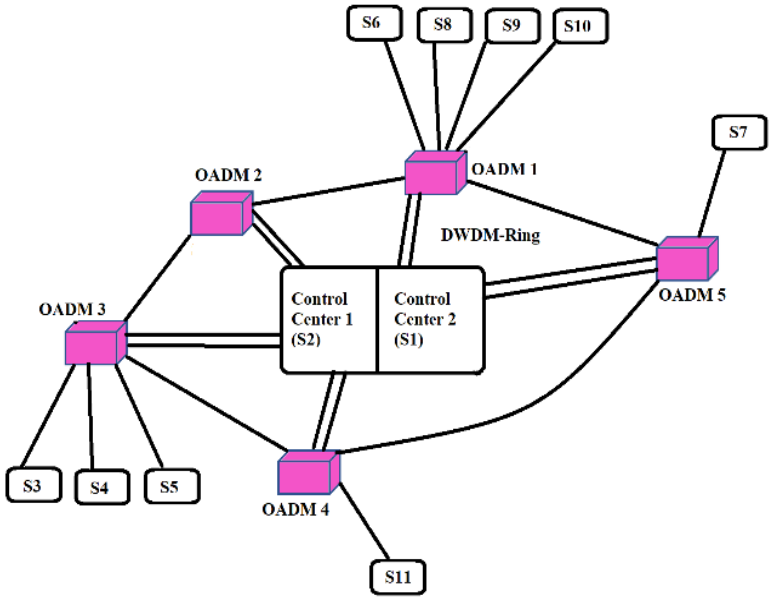

Fig. 3.4. DWDM-Ring Structure of IEEE 14-Bus System

3.1.2. Finding the Shortest Distance between All Pairs of Substations and Selection of Control Centers

In this step, the distance between a connected pair of substations is calculated first based on the length of the transmission line connecting them. The distance between all

pairs of substations is calculated next using the Floyd Warshall's all-pair shortest path algorithm [22]. Finally, two of the substations that are centrally located in the network and have large number of outgoing connections are selected as the primary and back-up control centers, respectively. For the IEEE 14-bus system, substation 2 is selected as the primary control center and substation 1 is selected as the secondary or backup control center (fig. 3.2.). Note that this step is needed for the realistic placement of the SONET and DWDM-Rings in a synthetic system, as elaborated in the subsequent steps. This step can be skipped if the locations of the SADMs, OADMs, and control centers are known in advance.

3.1.3. Placement of SADMs and Formation of SONET-Ring

SONETOE is a popular communication technology in which the SONET frames are directly carried on the Ethernet link layer. In this chapter, SONETOE technology [20] is used for transmitting RTU data from the substations to the control centers. SADMs are located in close proximity to the generating substations and the control centers as they are the most important substations of the system. Other substations transmit their SCADA data to the nearest SADM using the low bandwidth Ethernet channels. For fault tolerance all such SADMs are connected to each other via a ring structure, termed SONET-Ring. The link between two SADMs is bi-directional; therefore, even if a single link or node in the ring fails, the ring as a whole continues to function normally. In normal conditions, data from every SADM is sent to the control centers directly. However, if a link between a control center and an SADM fails, data from that SADM is forwarded to the next SADM in the ring which in turn forwards the data to the control centers.

In case of the IEEE 14-bus system, SADMs are placed near S2 (main control center), S1 (back up control center) and S3, S4, S5, and S10 (generating substations). Therefore, a total of six SADMs are placed in this system (see fig. 3.3.). Gateways of all other substations are connected to the nearest SADM in the network. Each substation is thus connected to an SADM, except, the control centers which are connected to all the SADMs in the ring. Fig. 3.3. shows the SONET-Ring structure of the IEEE 14-bus joint network. The control centers are placed in the center of the ring to show the star-ring topology of the network.

3.1.4. Placement of OADMs and Formation of DWDM-Ring

EoDWDM [21] is a low-cost high bandwidth technology that automates network management for better scalability and performance. This technology can automatically detect problems across the entire network and resolve them very fast. The proposed synthetic network uses this technology for the transfer of high-speed PMU data from the substations that have PMUs, to the control centers. Note that not all substations currently have PMUs. However, considering the steady growth in the field of ICT and the popularity of PMUs, the proposed design assumes that PMUs will be placed in every substation in the near future. Therefore, by default, every substation has a high bandwidth EoDWDM channel coming out of it and ending in the DWDM-Ring.

The DWDM-Ring of the proposed design is composed of low cost OADMs, each of which is placed near a substation having a PMU inside it. An OADM is placed near each of the control centers irrespective of whether it contains a PMU or not. Similar to the SONET-Ring, the DWDM-Ring is also bi-directional, ensuring fault tolerance. In case of

the IEEE 14-bus system, an OADM is placed near S2 (main control center), S1 (back up control center) and S4, S7, and S11 (PMU installed substations). Therefore, a total of five OADMs are placed in this system. Fig. 3.4. shows the DWDM-Ring structure of the IEEE 14-bus joint network.

3.2. Overview of MIIM and Modeling of IDRs

IDRs are logical equations that capture the interdependencies between two interacting entities. If they are written correctly, then by simply solving the IDRs after an entity or a set of entities have failed, the entities that will fail next can be identified. In this section, the creation of IDRs for the smart grid using MIIM is described. A smart grid system can be represented by the set $J\big(E, F(E)\big)$, where E is the set of entities in the joint network belonging to layers 1, 2, and 3 ($E = P \cup C \cup CP$) and $F(E)$ is the set of IDRs. In IIM [5], which was a precursor to MIIM, only structural dependencies were considered to formulate the IDRs. However, MIIM IDRs considers both the structural as well as the operational aspects of each of the entities during its formulation. In IIM, every entity was assigned a value of 0 or 1 depending on whether the entity was non-operational or operational. There was no concept of *reduced operability* in IIM, which is a common feature of most real systems. The entities in MIIM can take the following values: *0 indicating no operation, 1 indicating reduced operation*, and *2 indicating full operation*.

From an implementation viewpoint, IIM IDRs were composed of two standard Boolean operations, namely, AND (denoted by '.') and OR (denoted by '+'). In contrast, MIIM uses three new multi-valued operators for formulating the IDRs. The first operator is Operator 1, denoted by '○', which selects the lowest of its input values. The second operator

is Operator 2, denoted by '●', which selects the highest of its input values. The third operator is Operator 3, which is denoted by '◉'. If all the inputs of Operator 3 are same, then the output is also the same as the inputs; in all other cases the output is 1. The truth table for Operators 1, 2 and 3 are given in Table 3.1.

Input 1	**Input 2**	**Operator 1**	**Operator 2**	**Operator 3**
2	2	2	2	2
2	1	1	2	1
2	0	0	2	1
1	2	1	2	1
1	1	1	1	1
1	0	0	1	1
0	2	0	2	1
0	1	0	1	1
0	0	0	0	0

Table 3.1. Truth Table for MIIM Operators

Now the process of creating IDRs is illustrated using MIIM by deriving the IDRs for every entity of the IEEE 14-bus system. According to the design in fig. 3.3., every SADM is connected to its neighboring SADMs in the ring. In the MIIM IDR of $SADM_1$ ($C_{2,1,1,0}$) of the IEEE 14-bus system, this is expressed as:

$$C_{2,1,1,0} \leftarrow \left[\left(C_{2,1,2,0} \circ C_{2,2,1,2}\right) \bullet \left(C_{2,1,6,0} \circ C_{2,2,1,6}\right)\right] \equiv A \qquad (3.1)$$

Here, $C_{2,1,X,0}$ denotes $SADM_X$, while $C_{2,2,1,X}$ denotes the connection between $SADM_1$ and $SADM_X$, where X is the SADM ID. This IDR implies that $SADM_1$ remains operational if either the combination of $C_{2,1,2,0}$ AND $C_{2,2,1,2}$ is operational OR the combination of $C_{2,1,6,0}$ AND $C_{2,2,1,6}$ is operational. The SADMs can also forward the SCADA data collected from different substations to the control centers directly. Therefore, the MIIM IDR can be modified as:

$$C_{2,1,1,0} \leftarrow (A) \bullet \left[\left(C_{1,2,2,2} \circ C_{1,4,1,2}\right) \bullet \left(C_{1,2,1,1} \circ C_{1,4,1,1}\right)\right] \equiv B \quad (3.2)$$

Here, $C_{1,2,X,X}$ is the gateway of control center X and $C_{1,4,1,X}$ is the connection between $SADM_1$ and the gateway of X. Now, this $SADM_1$ also depends on the gateways of substations 2, 6, 7, 8, 9, and 11 for collecting SCADA data (see fig. 3.3.). Therefore, the IDR is updated as in eq 3.3.

The terms in blue denote the substation entities that provide SCADA data to this SADM. If all the gateways (2,6,7,8,9,11) from which $SADM_1$ receives SCADA data, remain operational then $SADM_1$ will work at its highest level of operation, i.e. 2. If one or more gateways fail or the connection between one such gateway and $SADM_1$ fails, then $SADM_1$ will work at a reduced level of operation, i.e. 1. If all the gateways connected to the SADM fails, then the SADM will also fail as it will have no data to carry to the control centers. Lastly, the SADM needs power supply to function. Hence, the IDR is further modified as in eq. 3.4.

$$C_{2,1,1,0} \leftarrow (B) \circ \Big(\left(C_{1,2,2,2} \circ C_{1,4,1,2}\right) \circledcirc \left(C_{1,2,6,6} \circ C_{1,4,1,6}\right) \circledcirc \left(C_{1,2,7,7} \circ C_{1,4,1,7}\right) \circledcirc \left(C_{1,2,8,8} \circ C_{1,4,1,8}\right) \circledcirc \left(C_{1,2,9,9} \circ C_{1,4,1,9}\right) \circledcirc \left(C_{1,2,11,11} \circ C_{1,4,1,11}\right)\Big) \equiv C \quad (3.3)$$

The final IDR of $SADM_1$ in eq 3.4 implies that it can receive power supply from any of the buses of any of the substations it is connected to; $P_4, P_7, P_9, P_5, P_6, P_{12}, P_{13}, P_{14}, P_{11}, P_{10}$ are the buses to which $SADM_1$ is connected and $L_{3,1}, L_{3,2}, L_{3,3}, L_{3,4}, L_{3,5}, L_{3,6}, L_{3,7}, L_{3,8}, L_{3,9}, L_{3,10}$ are the power supply lines to $SADM_1$. For

the $SADM_1$ to work, it should receive power from at least one of these buses. In this manner, the IDRs for all the six SADMs in the IEEE 14-bus system can be formulated. For creating the corresponding IIM IDRs, the '$\circ$' and '$\circledcirc$' operators must be replaced by '.' and the '$\bullet$' operator must be replaced by '+'. Similarly, the IDRs of OADMs can also be formulated for both MIIM and IIM.

$$SADM_1(C_{2,1,1,0}) \leftarrow (C) \circ [(P_4 \circ L_{3,1}) \bullet (P_7 \circ L_{3,2}) \bullet (P_9 \circ L_{3,3}) \bullet (P_5 \circ L_{3,4}) \bullet (P_6 \circ L_{3,5}) \bullet (P_{12} \circ L_{3,6}) \bullet (P_{13} \circ L_{3,7}) \bullet (P_{14} \circ L_{3,8}) \bullet (P_{11} \circ L_{3,9}) \bullet (P_{10} \circ L_{3,10})] \equiv D \quad (3.4)$$

Now, the IDR for the gateway of substation 1 can be formulated using the following set of steps:

Step 1: The substation gateway depends on the RTU of that substation for receiving SCADA data. This is described by,

$$C_{1,2,1,1} \leftarrow (R_1 \circ C_{1,6,1,1}) \equiv D \quad (3.5)$$

where $C_{1,2,1,1}$ is the gateway of substation 1, R_1 is the RTU of that substation and $C_{1,6,1,1}$ is the communication channel connecting the RTU to the gateway. If a substation has multiple RTUs, then the gateway of that substation collects data from all the RTUs of that substation.

Step 2: The substation gateway should also remain connected to at least one of the SADMs. It can receive SCADA data from other substations (if the gateway belongs to a control center) or it can send SCADA data to the control centers through the SONET-Ring. Also, if the gateway is connected to an SADM but the RTU of the substation does not work, then the gateway will not be able to send any data to the SADM. Finally, if one (or more in the case of control centers) SADM(s) connected to the gateway fail then the gateway will work at a reduced level of operation. This is described by the following IDR:

$$C_{1,2,1,1} \leftarrow \Big[(D) \circ \Big(\big(C_{2,1,1,0} \circ C_{1,4,1,1}\big) \circledcirc \big(C_{2,1,2,0} \circ C_{1,4,2,1}\big) \circledcirc \big(C_{2,1,3,0} \circ C_{1,4,3,1}\big) \circledcirc \big(C_{2,1,4,0} \circ C_{1,4,4,1}\big) \circledcirc \big(C_{2,1,5,0} \circ C_{1,4,5,1}\big) \circledcirc \big(C_{2,1,6,0} \circ C_{1,4,6,1}\big)\Big)\Big] \equiv E \quad (3.6)$$

In this IDR, $C_{2,1,X,0}$ is $SADM_X$ and $C_{1,4,X,Y}$ are the ethernet connections between $SADM_X$ and $Gateway_Y$.

Step 3: The substation gateway is also dependent on the PMU of that substation for receiving PMU data, i.e.

$$C_{1,2,1,1} \leftarrow \big(U_2 \circ C_{1,7,2,1}\big) \equiv F \quad (3.7)$$

In this IDR, U_2 is the PMU of that substation and $C_{1,7,2,1}$ is the communication channel connecting the PMU to the gateway. Similar to the case of RTUs, if a substation has multiple PMUs, then the gateway of that substation collects data from all the PMUs of that substation.

Step 4: The gateway should also remain connected to at least one OADM (similar to Step 2 in the case of SADMs). Hence,

$$C_{1,2,1,1} \leftarrow \Big[(F) \circ \Big(\big(C_{3,1,1,0} \circ C_{1,5,1,1}\big) \circledcirc \big(C_{3,1,2,0} \circ C_{1,5,2,1}\big) \circledcirc \big(C_{3,1,3,0} \circ C_{1,5,3,1}\big) \circledcirc \big(C_{3,1,4,0} \circ C_{1,5,4,1}\big) \circledcirc \big(C_{3,1,5,0} \circ C_{1,5,5,1}\big) \circledcirc \big(C_{3,1,6,0} \circ C_{1,5,6,1}\big)\Big)\Big] \equiv G \quad (3.8)$$

In the above IDR, $C_{3,1,X,0}$ is $OADM_X$ and $C_{1,5,X,Y}$ implies the DWDM connections between $OADM_X$ and $Gateway_Y$.

Step 5: The gateway should receive power from at least one of the buses in that substation.

Step 6: The gateway should remain connected to the substation server.

In order to obtain SCADA data from the buses of a substation, Steps 1, 2, 5, and 6 should be followed. Thus, the following IDR can be used to determine if a gateway is operational with respect to SCADA data. In other words, if the evaluation of the following IDR results in 2 (highest operational level) or 1 (reduced operational level), then the SCADA data from the corresponding buses can be received by the server of the substation.

The IDR in eq 3.9 is the final $Gateway_1^{SCADA}(C_{1,2,1,1})$ IDR for Case 1 where strictly separate channels are used for RTU and PMU data. However, for Case 2, if all the connections to the SADMs fail, the gateway can still receive SCADA data from the other substations if the data is sent through the high bandwidth EoDWDM network, i.e. through

the OADMs. Therefore, the above IDR can be further modified for Case 2 as shown below in eq. 3.10.

$$Gateway_1^{SCADA}(C_{1,2,1,1}) \leftarrow [C_{1,1,1,1} \circ C_{1,3,1,1}] \circ [E] \circ [(P_4 \circ L_{2,4}) \bullet (P_7 \circ L_{2,7}) \bullet (P_9 \circ L_{2,9}) \bullet (P_{Batt1} \circ L_{6,1})] \quad (3.9)$$

$$Gateway_1^{SCADA}(C_{1,2,1,1}) \leftarrow [C_{1,1,1,1} \circ C_{1,3,1,1}] \circ \Big[(D) \circ \Big(\big((C_{2,1,1,0} \circ C_{1,4,1,1}) \circledcirc (C_{2,1,2,0} \circ C_{1,4,2,1}) \circledcirc (C_{2,1,3,0} \circ C_{1,4,3,1}) \circledcirc (C_{2,1,4,0} \circ C_{1,4,4,1}) \circledcirc (C_{2,1,5,0} \circ C_{1,4,5,1}) \circledcirc (C_{2,1,6,0} \circ C_{1,4,6,1})\big) \bullet \big((C_{3,1,1,0} \circ C_{1,5,1,1}) \circledcirc (C_{3,1,2,0} \circ C_{1,5,2,1}) \circledcirc (C_{3,1,3,0} \circ C_{1,5,3,1}) \circledcirc (C_{3,1,4,0} \circ C_{1,5,4,1}) \circledcirc (C_{3,1,5,0} \circ C_{1,5,5,1}) \circledcirc (C_{3,1,6,0} \circ C_{1,5,6,1})\big)\Big)\Big] \circ [(P_4 \circ L_{2,4}) \bullet (P_7 \circ L_{2,7}) \bullet (P_9 \circ L_{2,9}) \bullet (P_{Batt1} \circ L_{6,1})] \quad (3.10)$$

In order to obtain PMU data from the buses of a substation, Steps 3, 4, 5, and 6 should be followed. The IDR in eq. 3.11. can be used to determine if a gateway is operational with respect to PMU data. In other words, if the evaluation of the following IDR results in

2 (highest operational level) or 1 (reduced operational level), then the PMU data from the corresponding buses can be received by the server of the substation.

$$Gateway_1^{PMU}(C_{1,2,1,1}) \leftarrow [C_{1,1,1,1} \circ C_{1,3,1,1}] \circ [G] \circ [(P_4 \circ L_{2,4}) \bullet (P_7 \circ L_{2,7}) \bullet (P_9 \circ L_{2,9}) \bullet (P_{Batt1} \circ L_{6,1})] \quad (3.11)$$

Now, gateway 1 is said to be fully operational if the following IDR gives a value 2, which implies that both SCADA and PMU data is sent (or received in case of control centers) by the gateway. If the IDR gives a value of 1, then it can be stated that either the PMU data or the SCADA data is sent/received by the gateway. If none of the two types of data is sent or received, then the evaluation of the following IDR will give 0.

$$C_{1,2,1,1} \leftarrow Gateway_1^{SCADA} \circledcirc Gateway_1^{PMU} \quad (3.12)$$

A substation server depends on the substation gateway and the power supply links from at least one of the buses of the substation. Therefore, the IDR of the server of substation 1 of IEEE 14-bus system can be written as:

$$C_{1,1,1,1} \leftarrow (C_{1,2,1,1} \circ C_{1,3,1,1}) \circ [(P_4 \circ L_{1,4}) \bullet (P_7 \circ L_{1,7}) \bullet (P_9 \circ L_{1,9}) \bullet (P_{Batt1} \circ L_{5,1})] \quad (3.13)$$

In the above IDR, $L_{1,4}, L_{1,7}, L_{1,9}$ are the power supply lines to the server from buses P_4, P_7, P_9, respectively. $L_{5,1}$ is the power supply line to the gateway from the battery backup P_{Batt1}. Following these steps, the IDRs of the substation servers and substation gateways for every substation can be derived for a synthetic joint network of a power system.

3.3. Case Study

The IEEE 14-bus network is used to illustrate the cascading failures that take place after a single failure occurs in the joint network. The failure which is simulated is a terrorist attack on substation 6 of this system. The physical attack leads to the immediate failure of Bus 12 (P_{12}), substation server ($C_{1,1,6,6}$), and substation gateway ($C_{1,2,6,6}$). The division of buses into substations for the IEEE 14-bus system is shown in fig. 3.2. The operational statuses of the communication entities which transfer the data to the control center are calculated using MIIM IDRs and IIM IDRs, respectively.

T1	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$				
T2	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$		
T3	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$	$C_{2,1,1,0} \rightarrow \mathbf{1}$	$C_{3,1,1,0} \rightarrow \mathbf{1}$

Table 3.2. Failure of Entities with Time obtained using MIIM

Table 3.2. shows how the smart grid system is affected gradually at each time step, denoted by *Ti*, if MIIM IDRs are employed. From Table 3.2., it is observed that as a result of substation 6 failure, bus P_{12}, gateway and server inside substation 6 fails immediately (at time instant T1). Consequently, the SONEToE and EoDWDM channels coming out of gateway 6 fail at the next time instant (T2). At T3, $SADM_1$ and $OADM_1$ start working at a reduced level of operation as they are not getting the expected data from gateway 6, but still get data from the other gateways to which they are connected. The cascading failure of entities stops at T3. The results obtained using MIIM IDRs are same irrespective of whether data transmission is done on the basis of Case 1 or Case 2.

T1	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$					
T2	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$			
T3	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$	$C_{2,1,1,0} \rightarrow \mathbf{0}$	$C_{3,1,1,0} \rightarrow \mathbf{0}$	
T4 CASE 1	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$	$C_{2,1,1,0} \rightarrow \mathbf{0}$	$C_{3,1,1,0} \rightarrow \mathbf{0}$	**NO SCADA** FROM $\boldsymbol{P_{10}, P_{11}, P_{13}, P_{14}}$
T4 CASE 2	$P_{12} \rightarrow 0$	$C_{1,1,6,6} \rightarrow 0$	$C_{1,2,6,6} \rightarrow 0$	$C_{1,4,1,6} \rightarrow 0$	$C_{1,5,1,6} \rightarrow 0$	$C_{2,1,1,0} \rightarrow \mathbf{0}$	$C_{3,1,1,0} \rightarrow \mathbf{0}$	**NO SCADA** FROM $\boldsymbol{P_{11}, P_{14}}$

Table 3.3. Failure of Entities with Time obtained using IIM

When IIM IDRs are used, two different results are obtained for the two cases of data transmission. Table III shows the cascading failure of entities obtained using IIM IDRs. The failure of entities at time instants T1 and T2 happen in the same way as in the case of MIIM. At T3, $SADM_1$ and $OADM_1$ fail completely due to the failure of one of the gateways (gateway 6) connected to them. This happens due to the binary nature of IIM, which does not account for reduced operability. Consequently, at time instant T4, no SCADA data is obtained from P_{10}, P_{11}, P_{13}, and P_{14} for Case 1, and P_{11} and P_{14} for Case 2. More entities fail in Case 1 than in Case 2 because in Case 2, unlike Case 1, the high bandwidth channel is capable of carrying both RTU and PMU data (Section 3.2.).

3.4. State Estimation Results

State estimation is performed for the IEEE 118-bus system to (1) understand if MIIM can predict the system state more accurately than IIM, and (2) demonstrate the scalability of MIIM. The state estimation is performed considering a single entity or

multiple entity failures and the states predicted using MIIM and IIM are both compared for the two different cases of communication discussed earlier.

3.4.1. Overview of State Estimation

The voltage magnitudes and angles (or the real and imaginary components of voltages) of all the buses constitute the states of the system. They are estimated using the formulated IDRs and the measurements obtained from the RTUs and PMUs. Note that loss of measurements from the sensors of a bus can result in a bad estimate of the state of that bus and/or the states of the neighboring buses. The relationship between the state matrix V and the measurement matrix Z for a bus that has a PMU placed on it, is given by:

$$Z = \begin{bmatrix} Z_r^S \\ Z_i^S \\ Z_r^P \\ Z_i^P \\ I_r \\ I_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \begin{bmatrix} V_r \\ V_i \end{bmatrix} \equiv JV \qquad (3.14)$$

where $Z_r^S,\ Z_i^S$ denote the real and imaginary voltages estimated using the traditional SCADA-based state estimation [23], $Z_r^P,\ Z_i^P$ denote the real and imaginary voltage measurements obtained from the PMU, and $I_r,\ I_i$ denote the real and imaginary (branch) current measurements obtained from the PMU. The matrices C_1, C_2, C_3, C_4 which relate the (branch) current measurements to the states of the system are obtained from the admittance matrix of the system. For instance, if a branch 'ab' with a series impedance $g_{ab} + j\ b_{ab}$ and shunt admittance $g_{a0} + j\ b_{a0}$ has a current I_{ab} flowing through it, then the relationship between the current in rectangular coordinates and the states of the system are given by:

$$\begin{bmatrix}(I_{ab})_r\\(I_{ab})_i\end{bmatrix} = \begin{bmatrix} g_{ab} & -g_{ab} & (-b_{ab}-b_{a0}) & b_{ab} \\ (b_{ab}+b_{a0}) & -b_{ab} & g_{ab} & -g_{ab}\end{bmatrix} \begin{bmatrix}(V_a)_r\\(V_a)_i\\(V_b)_r\\(V_b)_i\end{bmatrix} \quad (3.15)$$

The matrix J represents the matrix relating the measurements and the states of the system. V_r, V_i denotes the real and imaginary estimates of the states. In (eq. 3.15), the relation between the measurements and the states is linear, which means that it can be solved using the weighted least squares approach:

$$V = (J^T W^{-1} J)^{-1} ((J^T W^{-1})\, Z \quad (3.16)$$

In (3. 16), the matrix W is the final weight matrix comprising error covariance matrices of both SCADA and PMU measurements in rectangular form. The matrix W is obtained using the methodology developed in [24]. In this chapter, the SCADA measurement errors are assumed to be from a Gaussian distribution with 0 mean and 3% standard deviation, while the errors in the PMU measurements are assumed to be from a Gaussian distribution with 0 mean and 0.1% standard deviation.

3.4.2. Hardware Failure of Gateway 13 and SADM 39 of IEEE 118-Bus System

The gateway 13 is connected to bus 13 and SADM 39 is placed at substation 76 containing bus 85. Failure of gateway 13 results in loss of SCADA measurements at bus 14 for MIIM. The same original event results in the loss of SCADA measurements at buses 12, 13, 14, and 16, and loss of PMU measurements at bus 12 for IIM. The above-mentioned failures are common to both Case 1 and Case 2. Subsequent case-specific results are described below.

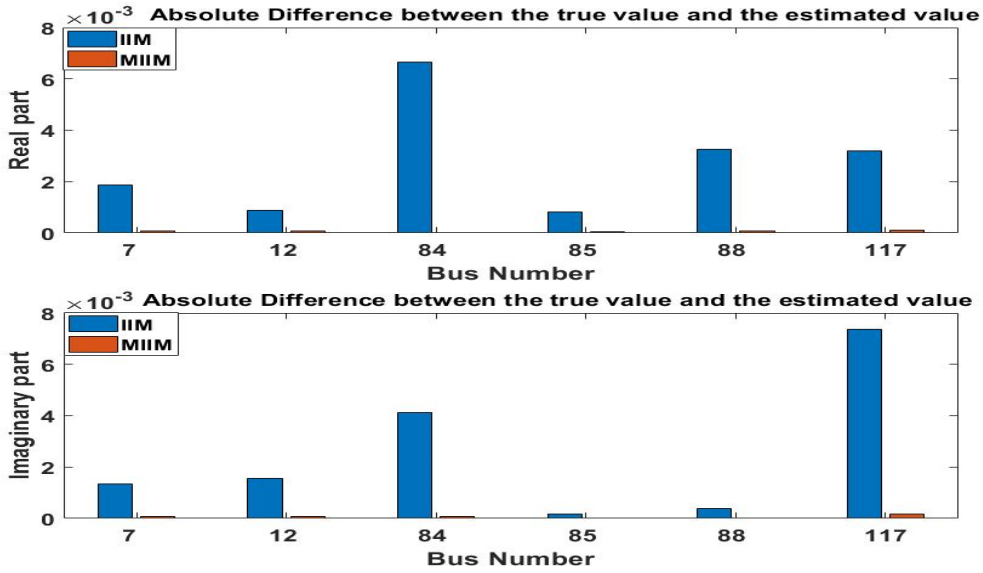

Fig. 3.5. State Estimation Result for Gateway 13 and SADM 39 Failure for Case 1

Case 1: As the high bandwidth channel cannot be used for carrying both RTU data and PMU data in this case, it results in an additional loss of SCADA measurements at buses 84, 85, and 88 for both MIIM and IIM, and a loss of PMU measurement at bus 85 for IIM due to SADM 39 failure. The state estimation results are shown in fig. 3.5, which depicts the absolute difference between the estimated value and the true value of the states for both the interdependency models. The buses 7, 12, 84, 85, 88, and 117 observe a significant difference between the estimated states for both the models. This is because, the buses 84 and 88 (7 and 117) are neighbors of bus 85 (eq. 3.12) which loses PMU data in the case of IIM, but not in the case of MIIM.

Case 2: In this case, the high bandwidth channel is capable of carrying both RTU and PMU data. Because of this, no subsequent failures take place for both IIM and MIIM. The results obtained on performing state estimation are shown in fig. 3.6. The difference

between the estimated states for both the models is considerable at buses 7, 12, and 117, due to the same reason mentioned in Case 1.

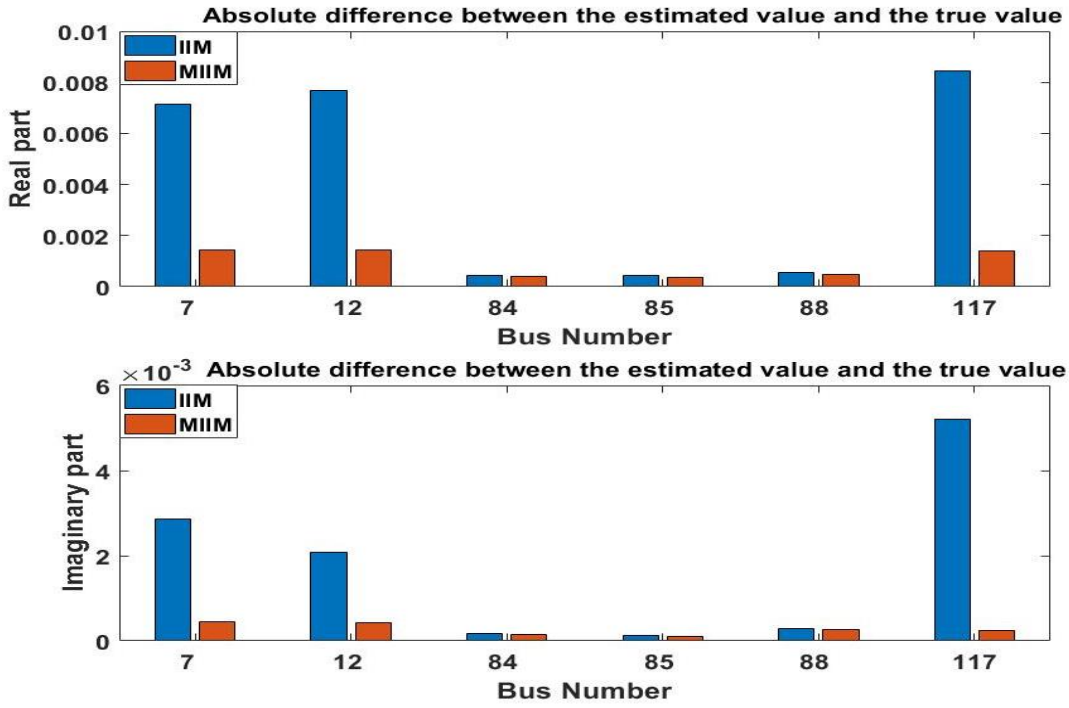

Fig. 3.6. State Estimation Result for Gateway 13 and SADM 39 Failure for Case 2

3.4.3. Damage of Substation 85 of IEEE 118-Bus System

Substation 85 consists of bus 95 of the IEEE 118-bus system. Damage to this substation would result in loss of all communication entities placed at or connected to substation 85. This results in PMU measurement losses at bus 94 and SCADA measurement losses at bus 94, 95, and 100 for IIM. However, it results in measurement loss at only bus 95 for MIIM. The above-mentioned failures are common to both Case 1 and Case 2. Subsequent case-specific results are described below.

Case 1: This case results in an additional loss of SCADA measurement at bus 101 for IIM. The state estimation results obtained during the substation 85 failure is shown in

fig. 3.7. A significant difference between the estimated states for both the models is observed at buses 93, 94, 95, and 100. This is because the buses 93, 95, and 100 are neighbors of bus 94 which loses PMU data in the case of IIM, but not in the case of MIIM.

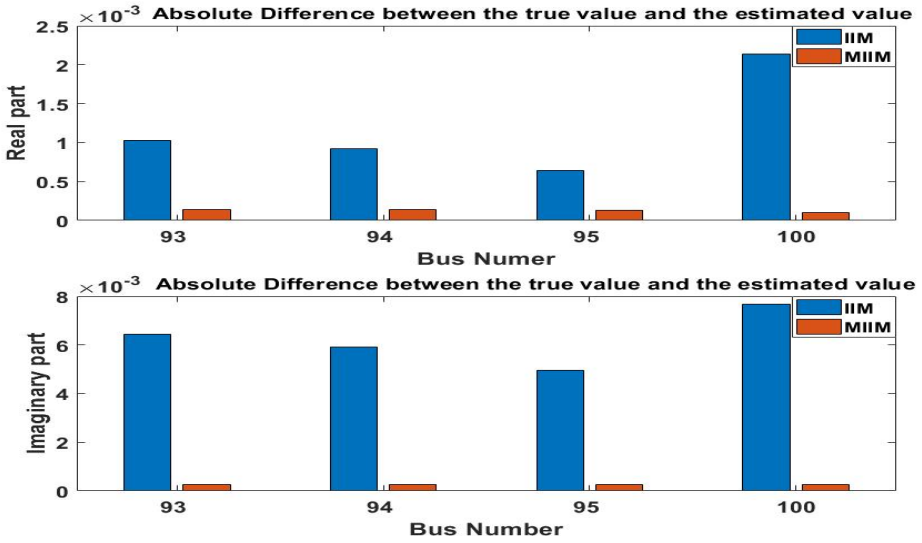

Fig. 3.7. State Estimation Result for Substation 85 Failure for Case 1

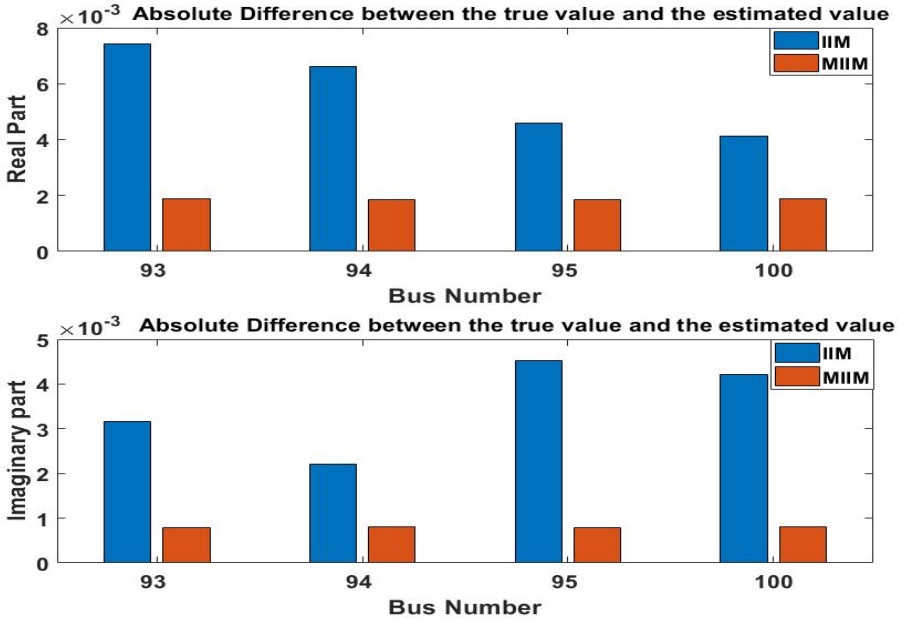

Fig. 3.8. State Estimation Result for Substation 85 Failure for Case 2

Case 2: In this case, since the high bandwidth channel is capable of carrying both RTU and PMU data, no subsequent failures take place for both IIM and MIIM. The results obtained on performing state estimation are shown in fig. 3.8. The buses 93, 94, 95, and 100 observe a notable difference between the estimated states for both the models for the same reason mentioned in the previous case.

The results obtained above confirm that the states of the system estimated using MIIM are closer to the true values than the ones obtained using IIM.

Chapter 4

IDENTIFICATION OF THE K-MOST VULNERABLE ENTITIES IN A SMART GRID

The smart grid system can be modelled as a joint power-communication network where the entities of both the network are structurally and functionally dependent on each other. The entities of communication network get power supplies from the power network entities and they in turn take the responsibility of monitoring the power network entities by continuously sending SCADA and PMU data from each of the substations to the control centers in a secure and efficient manner. As a result, entities of both power and communication network exhibit complex intra-and-interdependencies between them where the failure of one or more entities can lead to subsequent failure of multiple other entities leading to a catastrophe. In order to avoid such a condition and also to identify the most vulnerable entities in the system, the MIIM model [6] presented in Chapter 3 of this dissertation can be used. Just by solving the IDRs, the operators can identify the operational states of the smart grid entities. Therefore, at every instance all the vulnerabilities can be identified in the smart grid system.

Even after identifying all the vulnerable entities in the system, the smart grid operator can have a budget constraint of hardening only K entities of the network, where K can be any integer. In that case, it is important to identify the K-most critical entities in the system. The problem of identifying the K-most vulnerable entities in a joint power-communication network is already proved to be NP complete in [5]. Therefore, an ILP

based solution for the problem is given in this chapter using the MIIM IDRs. Finally, a validation of the results obtained from the proposed method is done by co-simulating the two layers of the smart grid network of IEEE 14-Bus system using MATPOWER and Java Network Simulator (JNS) and the simulation results are compared with that obtained using ILP solutions of both MIIM [6] and IIM [5] IDRs to show that the modified version of IIM is more realistic.

4.1. Problem Formulation

It is important for the operator of a smart grid to identify the most vulnerable entities in the network, even before any kind of failure or damage takes place in the system. An automated system offering identification of *K*-most Vulnerable Entities (KVE) in the steady state of a smart grid will help the operator to decide which of the entities in the system should be hardened [7], so that in any case the maximum damage in the smart grid can be avoided. When one or more entities fail in the smart grid system, many other entities fail as a result and this is called cascading failures, and this often might lead to a catastrophe if not arrested in time. This cascade stops when the system reaches a steady state once again.

Given an integer *K* and a power-communication network at a steady state, this problem returns the set of *K*-most critical entities in the joint network, failure of which can lead to the maximum total number of failed entities in the system at the end of the cascade propagation. It is to be noted that a cascade can only propagate in one direction since an already failed entity cannot be affected again by the cascading failure. Therefore, upper

bound of the cascade is $|EG|-1$; where EG is the total number of edges in the network. A formal definition of the KVE problem using the MIIM [6] model is as follows.

4.1.1. Inputs to the Problem

- (a) A joint network $J\big(E, F(E)\big)$; where $E = P \cup C \cup CP$.
    - $P = B \cup T \cup Batt$ (Buses, Transmission Lines/Transformers, Batteries)
    - $C = SE \cup SRE \cup DRE$ (Substation Entities, SONET-Ring Entities, DWDM-Ring Entities)
    - $CP = L \cup R \cup U$ (Power supply lines, Remote Terminal Units and Phasor Measurement Units)
- (b) Two positive integers *K* and *S.*

4.1.2. Decision Version of the Problem

Does there exist a set of *K* entities in *E* whose failure at time t would result in a partial failure of at least S entities or a complete failure of S/2 entities, in total at the end of the cascading process?

4.1.3. Optimization Version of the Problem

Compute the set of *K* entities in the joint network $J\big(E, F(E)\big)$ whose failure at time t would minimize the overall system state in the end of cascade propagation.

It is to be noted here, it is assumed that the set of *K* entities, failure of which minimizes the overall system state by causing either partial failure of *S* number of entities or total failure of $S/2$ number of entities as the most vulnerable set, where *S* is an integer less than or equal to the total number of entities in the smart grid.

4.2. Integer Linear Program based Optimal Solution

The problem of finding K-Contingency List is NP complete, which is already proved in [5] .Therefore, an ILP based solution for the problem is given in this chapter. Also, validation of the results should be done by comparing the ILP based solutions with the simulation results.

4.2.1. Variable List, Objective Function and Constraint Set

For each entity $e_i \in E$ a variable set $x_{i,t} \forall t, 0 \leq t \leq |E| - 1$ is created. The value of $x_{i,t}$ is 2 if it is fully operational, 1 if it is operating at a reduced level of operation and 0 if it is non-operational. The objective function for the problem can be defined as: $min \sum_{i=1}^{|E|} x_{i,|E|-1}$. This implies that, the problem aims at minimizing the system states for all the entities in the smart gird.

i. Constraint set 1: $\sum_{i=1}^{|E|} x_{i,0} = K$, entities failed at time step 0 is *K*.

ii. Constraint set 2: $x_{i,d} \leq x_{i,t-1}$, $\forall t, 1 \leq t \leq |E| - 1$. This implies that, an entity can only have a system state value at a time $t > d$, less than or equal to the system state value it had at time *d*.

iii. Constraint set 3: Based on the 3 new logical operations adopted by MIIM, IDRs can have the following format:$e_a \leftarrow (e_b \circledcirc e_c) \circ (e_m \bullet e_n)$.

a) Step 1: Firstly, the above IDR can be reformed as: $e_a \leftarrow z_{bcmn}$ where the new variable z_{bcmn} can be expressed as: $z_{bcmn} \leftarrow (g_{bc}) \circ (h_{mn})$ where the two new variables g_{bc} and h_{mn} can be further represented as: $g_{bc} \leftarrow e_b \circledcirc e_c$ and $h_{mn} \leftarrow e_m \bullet e_n$.

b) Step 2: Now, a linear constraint is developed for the z type variable (associated with Operator 1). In order to evaluate the IDR: $z_{bcmn} \leftarrow (g_{bc}) \circ (h_{mn})$, z_{bcmn} can be represented as: $z_{bcmn} \leq g_{bc,t-1}$ and $z_{bcmn} \leq h_{mn,t-1}, \forall t, 1 \leq t \leq |E| - 1$.

c) Step 3: A linear constraint is also developed for the h type variable (associated with Operator 2). In order to evaluate the IDR: $h_{mn} \leftarrow e_m \bullet e_n$, h_{pq} can be represented as: $h_{mn} \geq x_{m,t-1}$ and $h_{mn} \geq x_{n,t-1}, \forall t, 1 \leq t \leq |E| - 1$.

d) Step 4: For the g type variable, associated with the Operator 3, the following linear constraint is developed. The IDR: $g_{bc} \leftarrow e_b \circledcirc e_c$ is represented by the following set of linear equations: $g_{bc} \geq 0$, $g_{bc} \leq max\ \_state$, where max_state denotes the state value at the highest level of operability for an entity (2 in this case), and $N \times g_{bc} \leq x_{b,t-1} + x_{c,t-1}, \forall t, 1 \leq t \leq |E| - 1$.

In fig. 4.1., the maximum damage to the network after the failing of K-most vulnerable entities are predicted by the ILP based solution to the problem using MIIM IDRs and IIM IDRs. The predicted damages are compared with the simulated results for a smart grid system of IEEE-14Bus.

4.3. Comparative Analysis between IIM, MIIM and Simulation Results

In order to simulate the smart grid system with IEEE-14 Bus as the power network and a synthetic yet realistic communication network designed as per the principles proposed in [6], two different simulation platforms are selected. MATPOWER is selected for the simulation of the power layer whereas Java Network Simulator (JNS) is selected for the simulation of the communication layer. An event-driven synchronization [25] between these two kinds of simulation platforms is followed, in which whenever an entity in the power network fails at time $t = t'$, not only the power entities associated with that entity are updated but also all the communication entities receiving power from only that entity are removed before the starting of the next round of simulation at $t = (t' + 1)$. When a communication entity fails, it does not have any immediate effect on the power layer therefore only communication network is updated but eventually it will lead to unobservability of parts of the power layer even if does not lead to direct failure of other entities. However, this case is not considered in the simulation as it is assumed that by that time, either that communication entity will be replaced or repaired.

Identification of the K-most vulnerable entities using this setup is done by failing K-entities in the network at a time and observing the corresponding number of failed entities. Since this process is repeated for all different combinations of K-entities in the network to verify the results obtained from the ILP solutions using MIIM [6] and IIM [5] IDRs, a small smart grid system of only 14-Buses is considered here.

The same IEEE 14-Bus smart grid power-communication network as in Chapter 3 is considered for the comparative analysis in this chapter.

In fig. 4.1., the maximum damage to the network after the failing of K-most vulnerable entities are predicted by the ILP based solution to the problem using MIIM IDRs and IIM IDRs. The predicted damages are compared with the simulated results for a smart grid system of IEEE-14 Bus. The figure only shows the entities which are fully non-operational in case of MIIM and the simulation result, but IIM shows all entities which are assumed to be non-operational by IIM but some of them can actually be partially operational in reality.

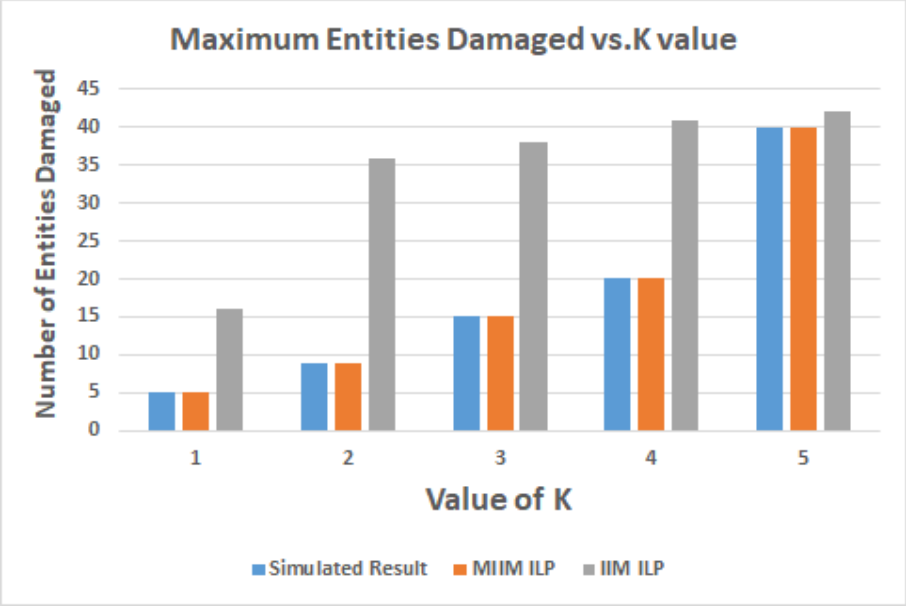

Fig. 4.1. Comparison between Simulation, MIIM ILP and IIM ILP Results

In the smart grid of IEEE 14-Bus system, there are 14 buses and 34 communication terminals like servers, gateways, SADMs and OADMs. It is considered that the transmission lines and communication channels can fail when the entities at the two ends of it also fail. Therefore, IDRs of those entities are considered. They can either have a state value 1 indicating they are operational or 0 denoting they have failed. However, the other 48 entities (14 *P type* and 34 *C type*) may depend on these transmission lines or

communication channels and thus they are included in the IDRs of those 48 entities. So, while finding the *K*-most vulnerable entities, only 48 entities are taken into account, but those 48 entities also cover the other entities which belong to categories like transmission lines or communication links. It is observed that for $K = 1$, the maximum damage in the network predicted by MIIM IDRs is 5 out of total 48 entities in the smart grid. According to the ILP based solution using MIIM IDRs, the most vulnerable entity in the network is Bus 7 denoted as P_7. Failure of P_7 leads to the failure of P_8 since it is connected to P_7 only. As a result, the communication entities in the substation of bus P_8 (substation 10 or S_{10}) also fail. Those entities are: S_{10} server denoted as ($C_{1,1,10,10}$) and S_{10} gateway denoted as ($C_{1,2,10,10}$). The $SADM_6$ or $(C_{2,1,6,0})$ is connected to S_{10} only, therefore it also fails. The simulation results also prove this. After the failure of P_7, S_{10} got islanded from the rest of the network; since all entities of that substation do not contribute to the smart grid anymore, they can be considered as non-operational. This observation can propose a theory that if the smart grid system is considered as a two layered graph, then the vertices in the power layer that are connected to some pendant vertices are the most critical for $K = 1$. However, if the IIM IDRs are used, the failure of any one of the following entities:

$$\left\{\begin{matrix}\{C_{1,1,6,6}\},\{C_{1,2,6,6}\},\{C_{1,1,7,7}\},\{C_{1,2,7,7}\},\{C_{1,1,8,8}\},\{C_{1,2,8,8}\},\\ \{C_{1,1,9,9}\},\{C_{1,2,9,9}\},\{C_{1,1,11,11}\},\{C_{1,2,11,11}\}\end{matrix}\right\}$$

will lead to the failure of $SADM_1$ and therefore all entities connected to $SADM_1$ *or* $(C_{2,1,1,0})$ will fail. This again will lead to the failure of $OADM_1, OADM_4$ *and* $OADM_5$. Also, due to the failure of $OADM_1$ *or* $(C_{3,1,1,0})$, communication entities of S_{10} also fail,

leading to the failure of $SADM_6$. The cascade of failure is shown below assuming substation server 6 or $(C_{1,1,6,6})$ failed initially:

$$\{C_{1,1,6,6}\} \rightarrow \{C_{1,2,6,6}\} \rightarrow \{C_{2,1,1,0}\} \rightarrow \left\{\{C_{1,2,7,7}\}, \{C_{1,2,8,8}\}, \{C_{1,2,9,9}\}, \{C_{1,2,11,11}\}\right\}$$
$$\rightarrow \left\{\{C_{1,1,7,7}\}, \{C_{1,1,8,8}\}, \{C_{1,1,9,9}\}, \{C_{1,1,11,11}\}\right\}$$
$$\rightarrow \left\{\{C_{3,1,1,0}\}, \{C_{3,1,4,0}\}, \{C_{3,1,5,0}\}\right\} \rightarrow \{C_{1,2,10,10}\} \rightarrow \{C_{1,1,10,10}\}$$

Therefore, a total of 16 entities are damaged. This prediction does not match with the simulation results.

For, $K = 2$, according to the MIIM IDRs, the failure of both the control center servers or failure of both control center gateways can result in the maximum damage i.e. all 34 communication entities will fail. However, this situation is absurd. Moreover, the main control center is already hardened by means of the backup control center. Analyzing the effect of failure of the backup entities as well (that also in the first step), is beyond the scope of this work and also it is hypothetical. Therefore, the maximum damage caused by the next 2-most critical entities in the network is being considered. The following pairs of entities are most critical if MIIM IDRs are used:

$$\begin{Bmatrix} \{P_1, P_7\}, \{P_2, P_4\}, \{P_2, P_5\}, \{P_2, P_7\}, \{P_3, P_7\}, \{C_{1,1,2,2}, P_7\}, \\ \{P_7, P_{10}\}, \{P_7, P_{13}\}, \{C_{1,2,2,2}, P_7\} \end{Bmatrix}$$

The failure of any of the above pairs can lead to the failure of a total of 9 entities in the network. The simulation results also validate this prediction. The cascading failure of entities after the initial failure of each of the above $(K = 2)$ sets is shown below:

$$Failure\ of\ Set\ 1: \{P_1, P_7\} \rightarrow \{P_1, P_7, P_8\} \rightarrow \left\{\{C_{1,2,3,3}\}, \{C_{1,1,3,3}\}, \{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}\right\}$$

$$\rightarrow \left\{\{C_{2,1,3,0}\}, \{C_{2,1,6,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 2: \{P_2, P_4\} \rightarrow \{P_2, P_4, P_3\} \rightarrow \{\{C_{1,2,4,4}\}, \{C_{1,1,4,4}\}, \{C_{1,2,5,5}\}, \{C_{1,1,5,5}\}\}$$

$$\rightarrow \left\{\{C_{2,1,4,0}\}, \{C_{2,1,5,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 3: \{P_2, P_5\} \rightarrow \{P_2, P_5, P_1\} \rightarrow \{\{C_{1,2,4,4}\}, \{C_{1,1,4,4}\}, \{C_{1,2,3,3}\}, \{C_{1,1,3,3}\}\}$$

$$\rightarrow \left\{\{C_{2,1,3,0}\}, \{C_{2,1,4,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 4: \{P_2, P_7\} \rightarrow \{P_2, P_7, P_8\} \rightarrow \left\{\{C_{1,2,4,4}\}, \{C_{1,1,4,4}\}, \{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}\right\}$$

$$\rightarrow \left\{\{C_{2,1,4,0}\}, \{C_{2,1,6,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 5: \{P_3, P_7\} \rightarrow \{P_3, P_7, P_8\} \rightarrow \left\{\{C_{1,2,5,5}\}, \{C_{1,1,5,5}\}, \{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}\right\}$$

$$\rightarrow \left\{\{C_{2,1,5,0}\}, \{C_{2,1,6,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 6: \{C_{1,1,2,2}, P_7\} \rightarrow \{\{C_{1,1,2,2}\}, \{P_7\}, \{C_{1,2,2,2}\}, \{P_8\}\}$$

$$\rightarrow \left\{\{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}\right\} \rightarrow \left\{\{C_{2,1,2,0}\}, \{C_{2,1,6,0}\}, \{C_{3,1,2,0}\}\right\}$$

$$= 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 7: \{P_7, P_{10}\} \rightarrow \{P_7, P_{10}, P_8\}$$

$$\rightarrow \left\{\{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}, \{C_{1,2,11,11}\}, \{C_{1,1,11,11}\}\right\} \rightarrow \left\{\{C_{3,1,4,0}\}, \{C_{2,1,6,0}\}\right\}$$

$$= 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 8: \{P_7, P_{13}\} \rightarrow \{P_7, P_{13}, P_8\} \rightarrow \left\{\{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}, \{C_{1,2,7,7}\}, \{C_{1,1,7,7}\}\right\}$$

$$\rightarrow \left\{\{C_{3,1,5,0}\}, \{C_{2,1,6,0}\}\right\} = 9\ Entities\ can\ fail\ as\ a\ result$$

$$Failure\ of\ Set\ 9: \{C_{1,2,2,2}, P_7\} \rightarrow \{\{C_{1,2,2,2}\}, \{P_7\}, \{C_{1,1,2,2}\}, \{P_8\}\}$$

$$\rightarrow \left\{\{C_{1,2,10,10}\}, \{C_{1,1,10,10}\}\right\} \rightarrow \left\{\{C_{2,1,2,0}\}, \{C_{2,1,6,0}\}, \{C_{3,1,2,0}\}\right\}$$

$$= 9\ Entities\ can\ fail\ as\ a\ result$$

When IIM IDRs are used, the following pairs become most critical entities:

$$\begin{Bmatrix} \{P_{12}, P_1\}, \{P_{13}, P_1\}, \{P_{14}, P_1\}, \{P_{11}, P_1\}, \{P_{10}, P_1\}, \{P_{12}, P_2\}, \\ \{P_{13}, P_2\}, \{P_{14}, P_2\}, \{P_{11}, P_2\}, \{P_{10}, P_2\}, \{P_{12}, P_3\}, \{P_{13}, P_3\}, \\ \{P_{14}, P_3\}, \{P_{11}, P_3\}, \{P_{10}, P_3\} \end{Bmatrix}$$

For each of the above pairs, the total damage will be: 36 entities. Considering the first ($K = 2$) set, the cascade will be:

$$\{P_{12}, P_1\} \rightarrow \{C_{1,2,6,6}, C_{1,2,3,3}\} \rightarrow \{C_{1,1,6,6}, C_{1,1,3,3}\} \rightarrow \{C_{2,1,1,0}, C_{2,1,3,0}, C_{3,1,3,0}\}$$

$$\rightarrow \{C_{1,2,7,7}, C_{1,2,8,8}, C_{1,2,9,9}, C_{1,2,11,11}, C_{1,2,4,4}, C_{1,2,5,5}\}$$

$$\rightarrow \{C_{1,1,7,7}, C_{1,1,8,8}, C_{1,1,9,9}, C_{1,1,11,11}, C_{1,1,4,4}, C_{1,1,5,5}\}$$

$$\rightarrow \{C_{2,1,4,0}, C_{2,1,5,0}, C_{3,1,1,0}, C_{3,1,4,0}, C_{3,1,5,0}\}$$

$$\rightarrow \{C_{1,2,10,10}, C_{1,1,10,10}, C_{3,1,2,0}, C_{1,2,1,1}, C_{1,1,1,1}\} \rightarrow \{C_{2,1,6,0}\} \rightarrow \{C_{2,1,2,0}\}$$

$$\rightarrow Total\ ICT\ network\ failure$$

In the similar way, prediction of MIIM based ILP solution is accurate in case of $K = 3, 4\ and\ 5$ but the prediction of IIM based solution is very different from the simulation results. It is observed that from $K = 4$ to $K = 5$, there is a sudden huge change in the total

number of failed entities for both MIIM based solution and simulation result. This happens due to the only set of $K = 5$, $\{P_4, P_7, P_9, P_5, P_6\}$ that leads to total communication failure in both the situations and as a result, the maximum total number of failed entities suddenly change from 20 $(for\ K = 4)$ to 40 $(for\ K = 5)$. The cascade of failure is shown below:

$$\{P_4, P_7, P_9, P_5, P_6\} \rightarrow \left\{\{C_{1,1,1,1}\}, \{C_{1,2,1,1}\}, \{C_{1,1,2,2}\}, \{C_{1,2,2,2}\}, \{P_8\}\right\}$$

which means both the control center servers and gateways are failing leading to total communication network failure. Therefore, total entities failed = (initial 5 buses + P8 + 34 communication entities) = 40.

Chapter 5

A SECURE SMART GRID MONITORING TECHNIQUE

Monitoring of smart grid system does not only signify keeping a track of the performance of its entities, rather observing the environment at the places where those entities are placed is also important. It should also be noted that damage to the smart grid entities is hugely caused by natural disasters like hurricanes, earthquakes, and forest fires. According to [26] since 1097 till today's date a large number of power grid substations have been damaged due to the most common natural disaster which is earthquake. Therefore, in this chapter, the main focus is on protecting the hubs of smart grid systems or the substations by pervasive monitoring of the seismic risks in the area around them while keeping track of the performance of smart grid entities by collecting information about the grid with the help of Phasor Measurement Units (PMUs).

Now, in order to provide continuous monitoring of the smart grid environment as well as entities, a secure and efficient communication network is necessary. In Chapter 3, a realistic design of the ICT system for smart grid is provided. However, that design relies completely on wired channels that either use SONET-over-Ethernet or Ethernet-over-Dense Wavelength Division Multiplexing. It is true that for transmitting the huge volume of data from smart sensor nodes like PMUs, wired communications can be the best choice. Yet, for sending data from environmental monitoring sensors like seismic sensors, a completely wired ICT system will neither be cost effective nor energy saving. Every communication entity in a wired network of chapter 3, draws power from the grid and thus

a huge amount of power is devoted for monitoring the power network itself. Moreover, addition or isolation of ICT entities for hardening purpose or fault tolerance, or during a failure, or a security threat is extremely difficult and costly in a wired system.

On the other hand, smart sensors like Phasor Measurement Units (PMUs) are already gaining popularity in smart grid system for measuring electrical waves. Power generation and transmission, power quality, equipment fitness, load capacity of equipment and load balancing in the grid can also be monitored by data sensing techniques. WSNs are comprised of low powered sensor nodes with easy installation process, lesser maintenance requirement, low installation cost, low power profile, high accuracy and scalability. All these have convinced the researchers [27] that WSNs are a very good choice for the designing of the ICT system of a smart grid. However, the most common drawback of a sensor node is that it is battery powered and it is not easy to replace its dead battery. As a result, energy conservation becomes important.

In this chapter, a new hybrid design of the ICT network for a smart grid is proposed using both wired and wireless communications. A WSN based communication network is used between seismic sensors or PMUs placed at the substations and a regional data aggregation point like a Regional Sink node (RS) for environmental data or Phasor Data Concentrator (PDC) for PMU data; and optical fiber-based communication is used between such regional aggregation point and the control centers. Energy efficiency of the WSN is obtained by following a two-fold method. Battery-powered relay nodes and an energy-efficient routing technique is followed for transmitting the seismic data from the substations to the control centers, while for sending the PMU data, the use of more

expensive rechargeable Energy Harvesting Relay Nodes (EHRNs) is done. It should also be mentioned that the sensor nodes and wireless channels are quite vulnerable to cyber-attacks. In this chapter, a Secure Smart Grid Monitoring Technique (SSGMT) [28] is proposed that aims at securing the sensed data by means of light weight security protocols used in [29] like Elliptic-Curve-Public-Key Cryptography (ECC), Elliptic-Curve-Diffie-Helman Key exchange scheme (ECDH), Nested Hash Message Authentication Codes (NHMAC) and RC5 symmetric cypher.

Now, in Chapter 3, an interdependency model, named as the Modified Implicative Interdependency Model (MIIM) is proposed that very accurately depicts the interdependencies between the different types of entities in the smart grid and also takes into account the interactions between them to determine the operational state of each such entity. MIIM uses multi-valued logic-based equations called Inter-Dependency Relations (IDRs) to model the structural and functional dependencies between entities in a smart grid. Just by solving such IDRs, the operational level of an entity can be identified. However, by solving IDRs it cannot be predicted if the data received from an operational entity is correct or it has false data injected into it. CCs completely depend on the data carried to it by the communication system from the PMUs to make all the required analysis. However, such data dependency is not covered in the dependency model MIIM. On the other hand, cyber-attacks like False Data Injections (FDI) are very common in the communication system of a smart grid. Cyber science researchers mainly focus on recovering the actual data [30] after a FDI attack takes place to perform state estimation. However, such processes are not only complex but also, they are time consuming. A better way of arresting

FDI attack can be identifying the attack before it reaches the CCs and thereby detecting the source of the attack, so that the source can be completely avoided for further communications. In this chapter, together with a hybrid communication network design an updated version of the MIIM is proposed and it is named as Multi-State Implicative Interdependency Model (MSIIM) which together with the structural and functional dependencies also takes into account the data dependency between the ICT entities of the smart grid. MSIIM ensures operational accuracy of the entities in the ICT network. A novel multi-path data routing technique is also proposed in this chapter which prevents FDI attacks and even if an attack takes place, that can be identified in the path to the CC and by solving the MSIIM IDRs, the source of the attack can be detected.

5.1. Overview of the ICT Network Setup Phase for SSGMT

In order to provide a reliable remote monitoring technique for the smart grid, a generic hybrid ICT system design is proposed in this section that can be applied on any given power network. In order to illustrate the steps of ICT network design, the generation and transmission part of a power grid formed by the IEEE 14-Bus system is considered. The network setup phase for SSGMT is divided into four steps and each step is explained with the help of an example power grid which is the IEEE 14-Bus system.

In the first step, a given power grid is divided into several substations. The buses connected by one or more transformers are placed in one substation with the assumption that transformers remain within a substation only [19] . This process is repeated until no transformers are left. Then each bus is placed in a unique substation. In fig. 5.1., the IEEE

14 bus system is divided into 11 substations denoted by S_i where i is the substation ID. After substation division the substation with maximum connectivity with other substations is selected as the main CC and that with the second highest connectivity is selected as the backup CC. After the CCs are selected, all substations are equipped with a router acting as a gateway (GW_i) for the substation S_i and the CCs are also equipped with servers ($Server_i$) which are access points for the operators. The CC-gateways can receive data from the RSs and the PDCs and send those data to server via a wired LAN connection in CCs. In the fig. 5.1., S_2 is the main CC and S_1 is the backup CC.

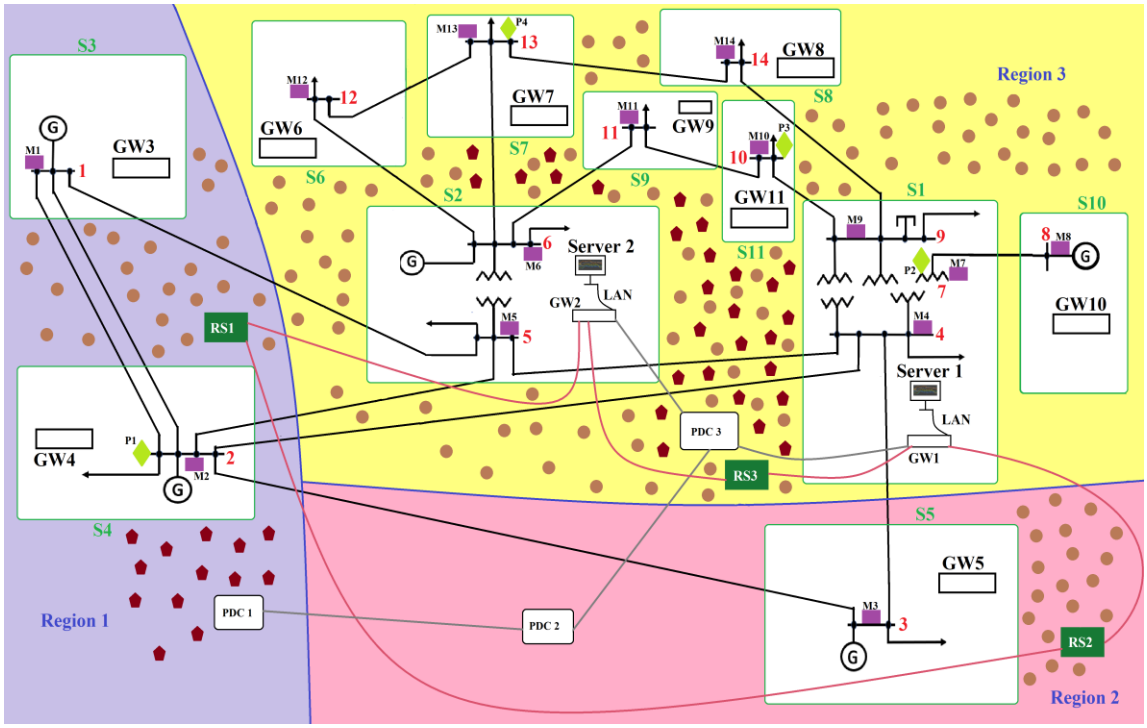

Fig. 5.1. Critical Information Infrastructure Design for a Smart Grid of IEEE 14-Bus

After the substation division, the distance between all pairs of substations $(S_i \ \& \ S_j)$ is calculated in the same way as in Chapter 3. Now, starting from a substation S_i with the maximum connectivity among the substations at the network borders, all substations which

are within a given distance of it, are marked as substations of a common monitoring region R_x. Then the next substation which is the closest to S_i but beyond the given distance and which is not yet placed in a monitoring region, is selected and the same process is repeated. This process is continued till every substation is placed within a monitoring region. For example, in the IEEE 14-Bus system, substations $S_3, S_4, S_5, S_6, S_7, S_8 \; and \; S_{10}$ are at the borders of the smart grid area. Among them, S_4 has the maximum connectivity, therefore the region division starts from it. Fig 5.1. shows that after this process is completed, the smart grid network for IEEE 14-Bus system is divided into 4 regions.

SSGMT considers two different types of Zigbee enabled smart sensors for monitoring the power network environment and the entities. The first type is the Measuring Unit (MU) based smart sensors [31]. The MU-based smart sensors have a sensing unit consisting of sensors like seismic sensors, heat sensors and wind motion sensors and a merging unit for performing signal amplification, conditioning and analog-digital conversion. There is an internal clock in the merging unit that can timestamp every sample and thereby time synchronize them with an external UTC (Coordinated Universal Time) source. The MU-based sensor has a network communication module which helps it to communicate with Intelligent Electronic Devices (IEDs) or applications. The second type of sensors used for SSGMT is the Zigbee enabled PMU-based smart sensors or ZPMU [32]. The sensing module of a ZPMU has sensors like voltage and current sensor which provide analog inputs for the PMU. Current and voltage phases are sampled at a rate of 48 samples per cycle in a PMU using a common time source for synchronization. It also has an internal clock and a processing module responsible for signal conditioning, timestamped

analog-to-digital signal conversion and computing voltage and current phasors to produce a synchronized phasor and frequency.

In this step, MU-based sensors (M_i) are placed at every substation but PMU-based sensors (PMU_i) are placed at some of the buses using an optimal PMU placement algorithm [vi]. $M_i s$ send data to the substation gateway (GW_i). Low-cost, non-rechargeable battery enabled relay nodes are randomly dispersed across the network area. These relay nodes can carry the environmental data to a RS placed at every monitoring region. Each RS is either connected to a neighboring RS or a CC-gateway via optical fiber channels to form a ring structure in order to provide fault tolerance. These RSs now convert the Zigbee signals obtained from the relay nodes to optical signals using a light emitting diode, associated with each RS. These optical signals are then carried to the CC-gateways via optical fiber channels and other RSs in the ring. TCP based communication is used between the RSs and the CC-gateway. It is to be noted that, the same network structure can also be used to transmit SCADA data from the substations to the CCs. However, SCADA uses fully wired communications since a long time and a shift in the network design may disturb the operation of the whole grid. That is the reason why new methods of transmitting SCADA data is not discussed here.

A phasor data concentrator (PDC), responsible for receiving and accumulating PMU data from multiple PMUs, is also placed at each region of the smart grid and few EHRNs are randomly deployed across the smart grid region. The idea behind the deployment of the two kinds of sensor nodes is that the cheaper non-rechargeable relay nodes will follow an event-driven hierarchical routing approach to send the environmental

data and the EHRNs will always be active to accumulate synchrophasor data from the substations of each region and send the data to the local PDCs and finally to the CC-gateways. Due to the high volume of PMU data transfer from each substation having a PMU, the sensor nodes carrying them should always be active. IEEE C37.118 standard is maintained for communication of PMU data to the PDCs. PDCs can convert the data to optical signals in the similar way as RSs and send that to the CC-gateways either directly or via other PDCs in neighboring regions. PDCs also use TCP based communication to send the optical data to CCs.

5.2. Risk Model and Assumptions for the SSGMT Network Setup

The WSN nodes are vulnerable to several attacks that are explored by researchers over the ages. In SSGMT, some of the common attacks on WSN are considered. The threat model for SSGMT predicts that a compromised entity of the ICT can pretend as a CC-gateway and congest the network with bogus communications to launch a flooding attack which can result in denial of service (DoS) by the ICT entities. Another common attack of the WSN is the Sybil attack [33] where a legitimate node of the network gets compromised and deceits the other nodes by providing multiple fake location information of itself. In SSGMT network setup, any ICT entity except the substation entities can impose a Sybil attack and oblige other entities to forward data to an entity that does not exist in that location resulting in increased packet drop. In a node compromise attack, the attacker physically gets hold of the node and reprograms it to launch attacks like wormhole attack and packet eavesdropping attack [33] . These attacks can also be launched by an outsider's node deployed in the network. The Sinkhole attack can be launched by attackers with high

quality equipment, and they can inform the ICT entities about a high-quality entity in their sensing zone, so that all the other entities select it to forward the data to the CC. This node can gather all the data and the CCs will not receive any information about the system state from the network.

In order to provide security to the ICT infrastructure for the smart grid, the following assumptions are made in the proposed work:

a) The substation equipment like the server and the gateway are trustworthy and it is impossible for an attacker to compromise those components. The servers authenticate themselves to the sensors or PDCs to avoid flooding attack [33] .
b) Each entity of the ICT network is provided with a global key GBK which an attacker cannot get hold of even if the entity is compromised.
c) A unique set of elliptic curves is stored in the memory of each ICT entity for the purpose of ECC and ECDH protocols. Also, in order to achieve those mechanisms, it is assumed that any pair of entities in the network agrees upon all the domain parameters of the elliptic curves stored in them.
d) The two CC-servers generate two separate one-way hash chains using SHA-1 hash function [29] . The starting key for each chain is randomly selected by the each of the servers. All the ICT entities are provided with same hash functions and the last keys of the two chains so that they can authenticate the control messages from the control center servers and avoid a Sinkhole attack.
e) Each ICT entity in the network is aware of its own ID, location information, region ID and ID of the servers.

f) All the PDCs can communicate with other PDCs in the neighboring regions.

5.3. SSGMT Routing Scheme

The goal of the ICT network for a smart grid is to securely transmit the sensed data from the sensors to the CCs and help in remote monitoring of the power grid. In order to achieve this with the help of a hybrid ICT network, the SSGMT is divided into 3 modules and described in this section.

5.3.1. Module 1: Data Forwarding to Substation Gateways by Sensors

In the first module of SSGMT, the M_i and PMU_i sensors placed in the substations sense environmental data and electrical waves from the buses they are placed on respectively and use Zigbee to send the data to the substation gateway GW_i. No security measure is adopted in this step as it is assumed that no cyber-attack can harm the communication within a substation. The CC-gateways forward the data directly to the connected servers, but they still need to send data to the other CC. So, all such gateways which need to send data outside their own substation, follow module 2.

5.3.2. Module 2: Data Forwarding by Substation Gateways to RSs and PDCs

The next phase of the hybrid ICT system of SSGMT is data forwarding by $GW_i s$. Each GW_i uses two separate methods for forwarding M_i and P_i data. First, the trust values (TV_i) of the non-rechargeable relay nodes (N_j) and EHRNs are determined by the $GW_i s$ of that region by means of forwarding a number of test messages (MSG) through them to

the RSs and PDCs of that region respectively. The format of the message is: $GW_i \rightarrow RS\ or\ PDC: MSG||K_x||HMAC(GBK; MSG||GW_iID||K_x)$. Here, the || symbol is used to denote concatenation or merging of different types of data in a data packet, K_x is a random key selected from the sequence of keys generated by each GW_i, using a one-way hash function. Once a random key K_x is used by a GW_i it is discarded for lifetime and never used again. A Hash based Message Authentication Code (HMAC) is generated over the MSG, gateway ID and the random key using the shared global key GBK and appended with the message, so that any non-legitimate node which do not have the GBK cannot separate the HMAC from the original message and therefore cannot overhear the message from the GW_i to the other legitimate ICT entities. The TVi of each N_j node and each EHRN is calculated using eq. 5.1. This TV_i of the nodes is recalculated at a regular interval.

$$TV_i = \left(\frac{MSG_{sent}}{MSG_{delivered}}\right) * 100 \tag{5.1}$$

Now, in order to send M_i data, the GW_is send a Forward_Request_Message (f_RQM) to each j^{th} non-rechargeable relay node (N_j) of that region at one-hop distance and a TV_i greater than 40% from the GW_i. The format of the request is: $GW_i \longrightarrow N_j: f\_RQM||GW_iID||HMAC(GBK; f\_RQM||GW_iID)$. Each N_j receiving the f_RQM not only sends back an acknowledgement (ACK) to the GW_i but also forwards the same f_RQM to the next hop nodes. They in turn send back ACK to the initiator GW_i following the same path. This process continues till nodes which are adjacent to the RS receive the f_RQM. In this process, each node can receive the f_RQM from different adjacent nodes but from the same GW_i and each time it will send back the ACK to the sender. This helps

the GW_i in verifying the consistency of response from the relay nodes. The format of the ACK is: $N_j \longrightarrow GW_i: ACK||BP_j||HMAC(GBK; ACK||BP_j)$. The ACK is appended with information like BP_j which stands for the remaining battery power of the node N_j.

After receiving ACK from all the N_j nodes, each GW_i selects a path to the RS of that region using a modified version of the Dijkstra's Shortest Path Algorithm. In this algorithm, the weight values W_{ij} of the link between two nodes i and j are determined using eq. 5.2.

$$W_{ij} = \frac{D_{ij}}{BP_j * TV_j} \tag{5.2}$$

Each GW_i now generates two distinct key pairs having a public and a private key $(pu1_{GW_i}, pv1_{GW_i})$ and $(pu2_{GW_i}, pv2_{GW_i})$ using ECDH and forwards the $pu1_{GW_i}$ to the RS of that region following the path created by the modified Dijkstra's algorithm and also forwards the $pu2_{GW_i}$ to the PDC of that region using a path with the most trusted EHRNs. The format of the message for sending the public key to the RS is given as: $GW_i \rightarrow RS: pu1_{GW_i}||GW_iID||HMAC(GBK; pu1_{GW_i}||GW_iID)$ and for sending the public key to the PDC the format is: $GW_i \rightarrow PDC: pu2_{GW_i}||GW_iID||HMAC(GBK; pu2_{GW_i}||GW_iID)$. The RS receiving the public key from GW_i also generates a private-public key pair (pv_{RS}, pu_{RS}) using ECHD and sends back its own public key pu_{RS} following the same path to GW_i. The gateway now computes a point in the elliptic curve $(x_k, y_k) = pv1_{GW_i}.pu_{RS}$. It is to be noted that that the public key of each entity is created using their own private key and the generator G of the elliptic curves stored in them. The RS also

computes a point $(x_k, y_k) = pu1_{GW_i}.pv_{RS}$ where x_k, the x coordinate of the computed point becomes the shared secret. This is ECDH key exchange scheme where the shared secret calculated by both parties is same– $pv1_{GW_i}.pu_{RS} = pv1_{GW_i}.pv_{RS}.G = pv_{RS}.pu1_{GW_i}$; where G is the generator of the elliptic curve. This shared secret x_k is the temporary key for exchanging encrypted data between the GW_i and RS. The same process is also followed between GW_i and PDC.

GW_i now encrypts the M_i data using the shared secret with RS and following RC5 symmetric cipher; and also, it encrypts the PMU_i data using the shared secret between itself and PDC following RC5 symmetric cipher. GW_i now forwards the encrypted M_i to RS and the encrypted PMU_i data to PDC following the paths used before. A nested HMAC is used by each gateway GW_i and the format of the message sent to RS is given as: $GW_i \rightarrow RS: EMD||GW_iID||HMAC(GBC; HMAC(x_k; EMD||GW_iID))$. Here EMD stands for the Encrypted MU-sensor data. Similarly, encrypted PMU_i data to PDC is forwarded in the format: $GW_i \rightarrow PDC: EPD||GW_iID||HMAC(GBC; HMAC(x_k; EPD||GW_iID))$. Here EPD stands for the Encrypted PMU-sensor data. The same route is followed for sending data until module 2 recalculates the TV_i values for the nodes. The RS and PDC of a region receive MU based sensor data and PMU based sensor data respectively from multiple GW_is.

5.3.3. Module 3: Data forwarding by RSs and PDCs to CC-gateways

In this module, the RSs and PDCs after obtaining the encrypted and HMAC-ed data from the N_js and EHRNs use the shared secret obtained for that sender GW_i to decrypt the

data packets. They also match the HMAC attached with the encrypted data to check if any false data injection took place. In case, the HMAC does not match, the data packet is dropped, and rerouting request is sent back to the sender. The main CC-server use ECC based public key cryptography and generate a public key for encryption and a private key for decryption of data. The ECC based public key of the main CC-server is sent to each of the RSs through the RS-ring and also to the PDCs via other PDCs and the optical channels. The main CC-gateway use a dedicated and secure optical channel to communicate with the backup CC-gateway. This channel is used to share the private key with the backup CC-server. RSs are responsible for data aggregation. Aggregated environmental data from the $N_j s$ are encrypted by the RSs using the public key of the main CC-server. This encrypted data is sent to both the CC-gateways via the RS-ring. In the similar way PDCs send the aggregated and encrypted synchrophasor data or PMU data via other PDCs to the CC-gateways wherefrom they reach the CC-servers.

5.4. Performance Evaluation and Simulation Results of SSGMT

Parameter	**Description**
Operating System	Fedora
Simulator	NS2.29
EHRN	NiMH
Battery capacity of EHRN	2000 (mAh)
Initial battery power of all nodes	150 (mAh)

Table 5.1. Parameter List for Simulation of SSGMT

In this section, the ICT network for a smart grid of IEEE 118-Bus system is considered. The total network region is divided into 8 regions and the power grid is divided into 107 substations. Substation 61 is selected as the main CC and it consists of 3 buses─68,

69 and 116. Substation 16, consisting of buses─17 and 30, is selected as the backup CC. In order to analyze the performance of SSGMT in this network setup, a total of 1500 non-rechargeable relay nodes, 500 EHRNs and 8 RSs and 8 PDCs are deployed in the network area and NS2.29 is used for simulation. The simulation results are compared with existing WSN based ICT systems for smart grid like Lo-ADI [34] and modified AODV [35] .

Fig. 5.2. shows the average communication delay calculated between the gateways and the CC for a given percentage of malicious nodes for all three ICT system designs−SSGMT, Lo-ADI and Modified AODV. It is evident from the figure that the delay in case of SSGMT is the minimum compared to the other two existing protocols. The main reason behind this is that SSGMT uses fast lightweight security mechanisms that can take place vey quickly and select the most trusted nodes to send data to the RS and PDC. Also, for sending data from the RS and PDC to the CCs SSGMT uses wired communication which is not only fast but more secure.

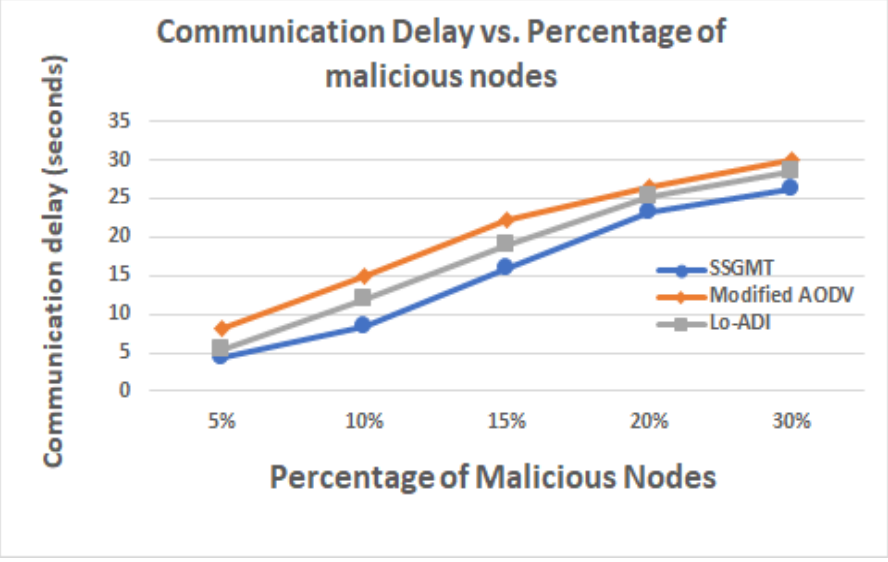

Fig. 5.2. Communication Delay vs. Malicious Nodes

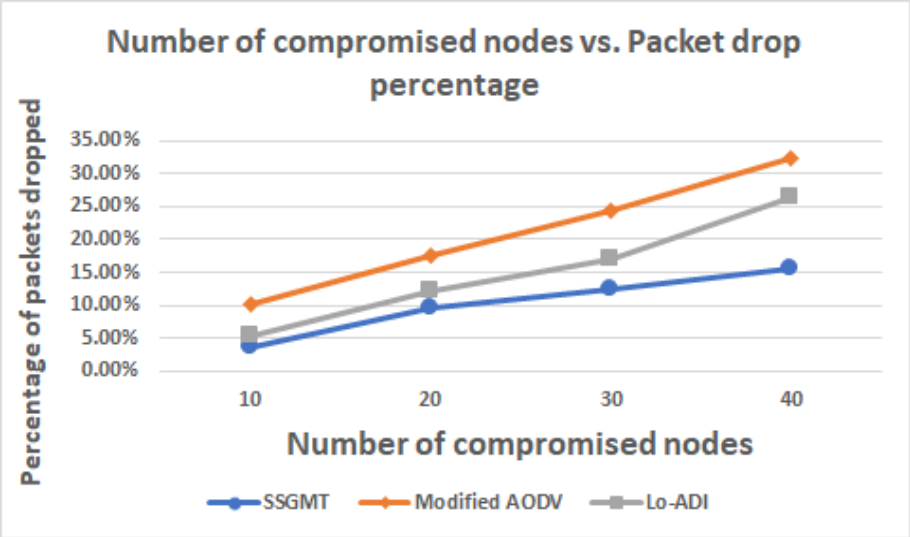

Fig. 5.3. Number of Compromised Nodes vs. Packet Drop

In fig. 5.3., it is proved that SSGMT also wins in terms of number of packets dropped when malicious nodes are present in the ICT system. Fig. 5.3. shows the percentage of packets dropped in presence of different number of malicious nodes for the three WSN based ICT networks. It is observed that the Lo-ADI algorithm performed better than the Modified AODV algorithm and the performance of SSGMT is better than both. Even in presence of 40 malicious nodes out of the total 2000 battery powered and EHRN nodes, that is when 2% of the total number of nodes is compromised, the percentage of packets dropped is 10% and it is to be noted that these dropped packets are mainly the data packets carrying MSG which are sent to evaluate the trust value of nodes.

5.5. Overview of the Multi State Implicative Interdependency Model (MSIIM)

In chapter 3, the smart grid is represented as a set $J(E, F(E))$ where $E = P \cup C \cup CP$ denoting the set of all entities in the smart grid and $F(E)$ the set of all IDRs. Similarly,

in this case, where MIIM is further updated to create the Multi State Implicative Interdependency Model (MSIIM), the same nomenclature is followed. All the power entities in both the models are denoted as P type entities where $P = \{P_1, P_2 \dots P_n\}$ and the communication entities are C type entities where $C = \{C_1, C_2 \dots C_m\}$. In MIIM, the set $F(E)$ only captures the structural and functional dependencies between the entities in the grid. In MSIIM, together with structural and functional or operational dependencies, the data dependency and operational accuracy of the entities is also taken into account while formulating the IDRs in the set $F(E)$. In MIIM, any entity can take a value of 0, 1 and 2; indicating no operation, reduced operation and full operation respectively. The novelty of MIIM lies in the fact that it considers a reduced operational level of the smart grid entities which is a very common feature of them. However, even if a communication entity is fully operational, it is not guaranteed to deliver correct data to an entity in the next-hop and this can be caused by attacks like node compromise attack or False Data Injection (FDI) attacks. This might lead to a wrong analysis of the system state by the CCs and thereby compel the operator to take wrong decision. In MSIIM, the P type entities can take the same 3 states as described in chapter 3, but the state values are− 3 for full operation, 2 for reduced operation and 0 for no-operation. This is done to have a consistency in the operational levels with the C and CP type entities yet, the P type entities can never take a state value of 1, as of now. Yet, this is a scope of the future work where the operational levels for the P type entities can be increased. The C and CP type entities can take four different states in MSIIM as the reduced operation level in chapter 3, is further categorized into two types here, namely−operation reduced by false data and operation reduced by interdependent

non-operational entities. Therefore, the four different states of *C* and *CP* type entities in MSIIM are denoted by─ 0 indicating no-operation, 1 indicating reduced operation by false data, 2 indicating reduced operation by interdependent non-operational entities and 3 indicating full operation.

In MSIIM the same three new logical operators are used as in chapter 3, except the inputs and outputs can range from 0 to 3 in place of 0 to 2. The truth tables for the 3 operators of MSIIM are given in Table 5.2.

Input 1	**Input 2**	**Operator 1 (●)**	**Operator 2 (○)**	**Operator 3 (◉)**
3	3	3	3	3
3	2	3	2	2
3	1	3	1	1
3	0	3	0	2
2	2	2	2	2
2	1	2	1	1
2	0	2	0	2
1	1	1	1	1
1	0	1	0	1
0	0	0	0	0

Table 5.2. Truth Table for MSIIM Operators

	MIIM	**MSIIM**
STEP 1	$C_l \rightarrow 1$	$C_l \rightarrow 2$
STEP 2	$C_i \leftarrow \big((3 \circ 3)\bullet(3 \circ 3)\big)\circledcirc 1\big)$	$C_i \leftarrow \big((2 \circ 2)\bullet(2 \circ 2)\big)\circledcirc 2\big)$
STEP 3	$C_i \leftarrow ((3\bullet 3)\circledcirc 1)$	$C_i \leftarrow ((2\bullet 2)\circledcirc 2)$
STEP 4	$C_i \leftarrow (3\circledcirc 1)$	$C_i \leftarrow (2\circledcirc 2)$
STEP 5	$C_i \leftarrow 1$	$C_i \leftarrow 2$

Table 5.3. Evaluation of IDRs to Obtain State Values

In order to explain the difference between the two dependency models, a set of dependencies between C type and P type entities are considered as follows. C_i, a C type,

be operational if (i) C_j which is another communication entity and P_a which is a P type entity, are operational, or (ii) C_k which is another C type entity and P_b which is another P type entity, are operational, and (iii) C_l which is another C type entity is operational. Now, it is assumed that even if an entity in condition (i) or (ii) fails, C_i will still work with full operability, but if (iii) is not satisfied then C_i will operate at a reduced level; this can be expressed using MIIM or MSIIM as: $C_i \leftarrow \Big(\big(C_j \circ P_a\big) \bullet (C_k \circ P_b)\Big) \circledcirc C_l$.

In the first step of the Table 5.3. itself, the false data injection is not captured by MIIM as the entity is fully operational other than that. MIIM is not designed to capture data dependencies. So, even after evaluating the IDRs, this fault of the entity C_l is not reflected in the state value of C_i which depends on the data from C_l. However, in MSIIM, all the entities getting data from C_l will be having a state value of 1 and this will help the destination node or a PDC to trace back through the path of entities having a state value 1 and identify the entity injecting false data into the system. Yet, the challenge lies in detecting the false data injection by a particular node or ICT channel. In this chapter, it is assumed that a node injecting false data into the system cannot identify that, but all nodes receiving data from that compromised node can identify the false data injection using the technique described in section 5.6. and thereby, they update the state value of the node in the previous hop to 1. Unlike MIIM IDRs which are evaluated by the CCs based on data received from the RTUs and PMUs, each communication entity of the smart grid in MSIIM can evaluate their own IDR and send the state value to the CC. Moreover, IDRs are re-generated for only the communication entities which are selected for sending data to the CC in that round of data routing. Therefore, the IDR for a *C* type entity consists of other *C*

type entities which are also selected for that round. Only the IDRs of *P* type entities are evaluated by the CCs. A distributed state table is maintained in MSIIM where each entity has the state values of all its connected entities. Each time data is received from some of the *C* type entities connected to it and also when data is successfully sent to the next hop, the state table is updated. In case of a false data injection, the compromised node will have a state value of 2 or 3 and all its neighboring C type entities getting data from it will have its state value as 1. This will help the PDC to identify the compromised node. Section 5.6. illustrates this with the help of fig. 5.5.

5.6. Routing of PMU data from Substations to Control Centers Using MSIIM

In this section, only the part network responsible for transmitting the PMU data from substations to CCs, described in section 5.1. is taken into account for discussing the operation of the MSIIM model. However, the same technique can be applied to any other communication network as well for identifying FDI attacks.

5.6.1. Assumptions for Designing the Secure Routing Scheme using MSIIM

In order to design the secure routing scheme with the MSIIM model, the following assumptions are made:

- PMUs, CC-gateways and CC-servers are trusted.
- FDI attack can take place at any point after the data is sent by the PMUs and before it is received by the CC-gateways.

- An MSIIM state table is maintained by each communication entity in the smart grid and whenever, an entity gets involved in the routing process, it recalculates its own state value based on the state values of entities connected to it and the MSIIM IDRs.

5.6.2. Secure Routing Scheme using MSIIM

The MSIIM based secure routing technique for a smart grid system is divided into three modules as follows:

5.6.2.1. Module 1: Data Forwarding to Substation Gateways by PMUs

The PMUs generate 48 time-stamped samples of data per processing unit clock cycle from the analog signals received from the current and voltage sensors. The samples are sent to the $GW_i s$ in which the PMUs are placed, using a dedicated wireless communication channel. Although PMUs can also be compromised to launch different attacks, in this thesis, PMUs are considered trusted and PMU compromise is kept as a scope of future works. Data from the PMUs of a single substation are stored in a data queue by the gateway of that substation GW_i . The MSIIM IDR for a substation gateway is given as: $GW_i \leftarrow [(PMU_1 ◉ PMU_2 \dots ◉ PMU_n) \circ (EHRN_1 ◉ EHRN_2 \dots ◉ EHRN_m)]$ assuming there are n number of PMUs in the substation and m number of energy harvesting relay nodes at one-hop distance from the GW_i. If the GW_i receives data from all the PMUs of that substation, then it works at operation level equal to 3. If any one of the PMUs don't operate, then its state changes to 2. Similarly, it should remain connected to all the relay nodes at one hop distance and at least one of those EHRNs should be operating at a level greater

than 1 to keep GW_i operational. After calculating its own state value, the GW_i starts Module 2 of the secure routing technique.

5.6.2.2. Module 2: Data Forwarding by Substation Gateways to PDCs

In this module, each gateway GW_i discovers all possible paths from itself to the PDC of that region, consisting of nodes having a state value ≥ 1. GW_i calculates the trust values TV_i of each path by forwarding a number of test messages through them. TV_i of the i^{th} path is calculated using equation 5.3.

$$TV_i = \sum_{j=1}^{Number\ of\ nodes\ in\ path\ i} \frac{test\ messages\ delivered\ by\ node\ j}{test\ messages\ sent\ to\ node\ j} \qquad (5.3)$$

GW_i then selects four paths with the highest TV_i. Pair-wise keys (PK_{xy}) are generated using Diffie-Hellman Key exchange scheme and are traded between every adjacent pair of nodes x-y in each of those four paths and also the MSIIM IDR of each node in a path is also generated by each node. The format of MSIIM IDR for an EHRN in the trusted path will depend on the ICT entities adjacent to that particular entity which are included in the four trusted paths selected by the GW_i. GW_i now generates a secret key SK_i and shares it with the PDC of that region in the following way:

- Four random equal sized binary numbers–$Key_1, Key_2, Key_3, and\ Key_4$ are generated by GW_i.
- A secret key (SK_i) is now generated by GW_i by XORing the four binary numbers.

$$SK_i = Key_1\ XOR\ Key_2\ XOR\ Key_3\ XOR\ Key_4 \qquad (5.4)$$

- Now, each such binary number Key_i is XORed with a binary number with equal number of 1s to generate a key fragment in the following way. Let Key_i be of size 4 bits. $Key_i = B_1B_2B_3B_4$ where B_i is a binary digit or a bit. Key Fragment (KF_i) is generated as: $KF_i = Key_i\ XOR\ 1111$.
- Now, each of 4 key fragments is encrypted with RC5 symmetric cipher using the pairwise key that the gateway shares with each of the next-hop nodes in the four most trusted paths. A Hashed Message Authentication Code (HMAC) [29] is generated over the total message using the same pairwise key. The format of the message is: $GW_i \rightarrow *: KF_i||GW_iID||HMAC(KF_i||GW_iID)$. Here * stands for all the nodes at one-hop distance from GW_i which are in the selected trusted paths and KF_i is the i^{th} key fragment.
- Each node in the trusted path matches the HMAC using the same pairwise key that it shares with the previous node in the path and if match is found, a new HMAC is generated over the encrypted key fragment using the pairwise key of the next node and this process is repeated till the encrypted key fragment reaches the PDC.
- The PDC receives the four key fragments via the four trusted paths and re-generates the secret key in the following way:

$$SK_i = KF_1\ XOR\ KF_2\ XOR\ KF_3\ XOR\ KF_4 \quad (5.5)$$

The key fragment sent via i^{th} path is stored by the nodes in that path. Any attacker which even if successfully gets hold of a key fragment they can never regenerate the secret key without having all the fragments.

The GW_i now receives 48 samples from the PMUs every second and these 48 samples are divided into 4 parts with 12 samples in each part. Each of these four parts of PMU data is now encrypted by the GW_i with RC5 symmetric cipher using the secret key and a HMAC is generated using the same secret key over the encrypted data. Four different trusted paths are selected to send the encrypted samples to the PDCs. A nested HMAC (NHMAC) is generated over the whole data using the KF_i send through that path and another NHMAC is generated using the pairwise key shared between the next-hop nodes in each of the trusted paths. The encrypted and HMAC-ed data is now forwarded to each of the four trusted paths.

The format of the data fragment sent via the i^{th} path is: $GW_i \rightarrow R_i: EDF||SV_{GW_i}||HMAC(PK_i; HMAC(KF_i; HMAC(SK_i; EDF||SV_{GW_i})))$ where EDF is the encrypted data fragment and SV is the state value of the sender. This process is repeated for all four parts of the 48 samples from a PMU and they are forwarded to the PDC using the four most trusted paths. The same paths are used by the GW_i to forward data to PDC until an FDI attack is detected in one of the paths. The nested HMAC using PK_i helps in identifying FDI attacks by a compromised communication channel between two nodes. The nested HMAC using KF_i helps in identifying the FDI attack by any compromised node and any unobservable attack can be finally detected by the PDC using the HMAC over the encrypted data generated using the secret key SK_i.

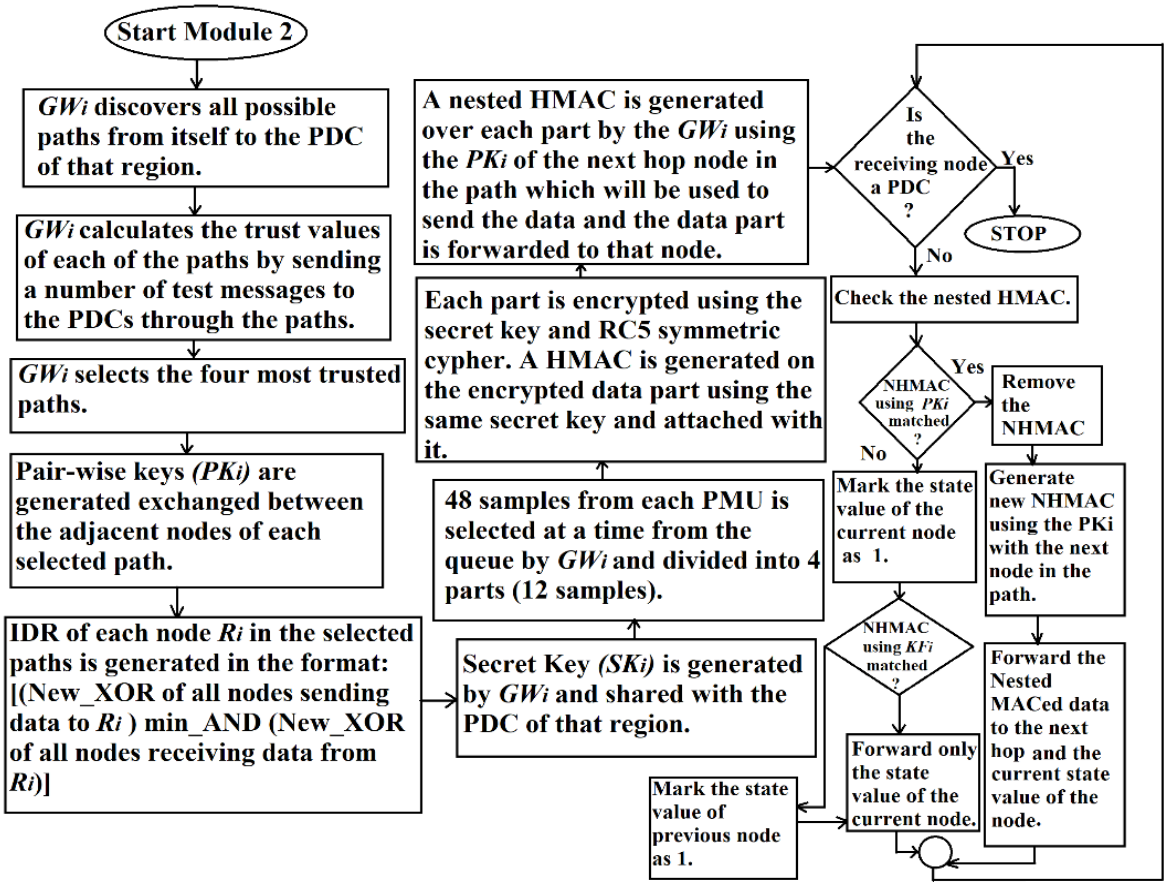

Fig. 5.4. Flowchart Describing Module 2 of Secure Routing Scheme

Each relay node R_i receiving a data fragment from the previous node in the i^{th} trusted path, first matches the HMAC generated using its own pairwise key with the previous-hop node. If match is not found, it marks its own state value as 1. It stops forwarding the data fragment and sends its own state value to the next node. The next node calculates its own state value using MSIIM IDR and forwards the same to its next hop, and finally it is forwarded to the PDC. The PDC now uses the state values of the nodes to trace the node injecting false data [fig. 5.5.].

In fig. 5.5. (a), the trusted paths are: $\{R_1 \rightarrow R_4 \rightarrow R_7\}$, $\{R_3 \rightarrow R_7\}$, $\{R_6 \rightarrow R_{10} \rightarrow R_{12}\}$ and $\{R_6 \rightarrow R_9 \rightarrow R_{11} \rightarrow R_{12}\}$. In fig. 5.5. (b), relay node R_9 is compromised and it is injecting false data into the system, which is first noticed by R_{11} and it only forwards the state value to the next node R_{12} and it also calculates its own state value in fig. 5.5. (c). R_{11} also changes the state value of R_9 to 1, in its own state table. The compromised R_9 will still maintain a state value greater than 1 in its own state table. PDC_1 checks the state value of R_{12} and it is 1 in both its own table and the state table of R_{12}. PDC_1 now checks the next-hop nodes connected to R_{12}. R_{10} has a state value 3 and it is same in the state table of R_{12} as well. Then that path is not checked any more as it is recognized as a trusted path without having any inconsistencies in the state values of the nodes. PDC_1 finds the state value of R_{11} and R_9 as 1. When the state table of R_9 is checked, PDC_1 finds that it has a state value greater than 1. Now R_9 is identified as the node injecting false data and the PDC immediately sends the ID of R_9 to both the CCs, so that R_9 is removed from the list of nodes in the ICT network of the smart grid. The PDC also sends the ID of R_9 and a Negative Acknowledgement (NAK) to GW_i to inform it did not receive the data packet sent via R_9. GW_i discards all paths containing R_9 and selects another path with highest trust value from its stored list of discovered paths to resend that data fragment to PDC. If no path is left in the discovered set of paths, then module 2 is repeated from the beginning.

In case a FDI attack is not detected by the relay nodes of a path but the NHMAC using the SK_i generated by the PDC does not match with the attached NHMAC, then the whole path is marked as unsafe and IDs of all nodes in that path are sent to the CCs. Also,

a NAK is sent to the GW_i mentioning the IDs of the path marked as unsafe, so that all paths having those nodes are removed from the discovered set.

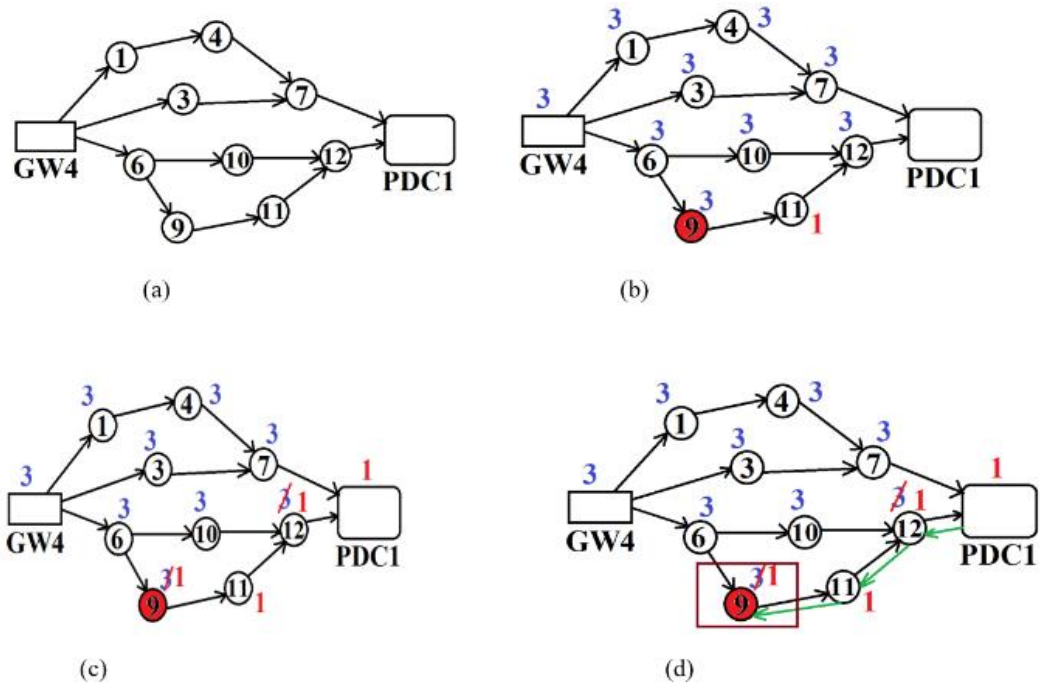

Fig. 5.5. Identification of FDI Attack by PDC

5.6.2.3. Module 3: Data Forwarding by PDCs to CC-gateways

Each PDC after receiving data from all the four paths decrypt them using the secret key. Same process is repeated for data obtained from all PMU containing substations. PDC aggregates the data from all the PMUs. Each of the CC-gateways use Elliptic Curve Diffie-Helman (ECDH) [29] key exchange scheme to establish a shared secret key with the PDC. The GW_i uses this secret key to encrypt the aggregated data and it also generates a HMAC over the encrypted data using the same shared secret key and attach it with the aggregated. The encrypted and aggregated data is now sent to the respective CC-gateways through the dedicated fiber optic cables connecting the PDC to each of the CC-gateways.

5.7. Performance Analysis of MSIIM Simulation Results for the Routing Scheme

Parameter	**Description**
Operating system	Fedora
Simulator	NS2.29
Rechargeable battery for EHSNs	NiMH
Communication standard	Zigbee
Battery capacity of EHSNs	2000(mAh)
Initial battery power for all nodes	200 (mAh)

Table 5.4. Parameter List for Simulation of the Routing Scheme using MSIIM

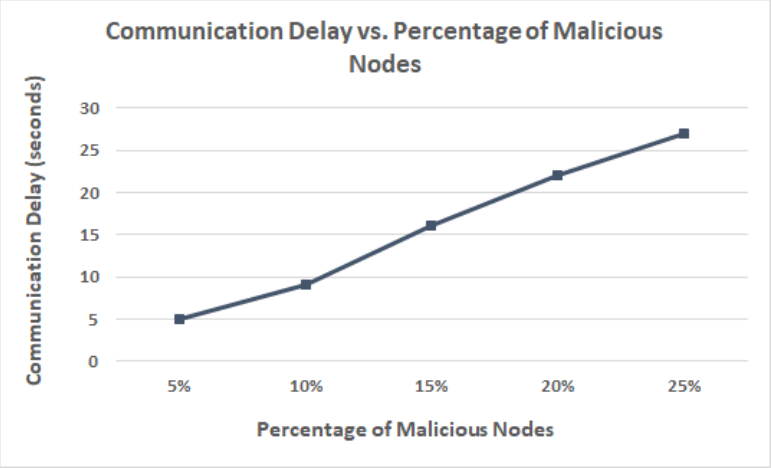

Fig. 5.6. Percentage of Node Compromise vs. Communication Delay

In order to analyze the performance of the proposed secure routing technique that relies on the MSIIM model, a smart grid system of IEEE 118-Bus is considered. The network area is divided into 8 regions and there are 5 PDCs. A total of 200 EHRNs are deployed over the network region. The power grid is divided into 107 substations; S_{61} having buses $P_{68}, P_{69}\ and\ P_{116}$ is selected as the main CC and S_{16} having buses P_{17} and P_{30} is selected as the backup CC. The simulation platform and parameter list are given in Table 5.4.

In fig. 5.6., the communication delay vs. the percentage of node compromise is shown. It is observed that the communication delay suddenly increases when the percentage of malicious nodes increases from 10 to 15 but it drops again when the percentage increases from 15 to 20. When a certain number of malicious nodes are detected in the network region and the substation gateways have to rediscover new set of paths to the PDC, the communication delay increases. Again, when a number of paths are discovered and stored, increase in the number of malicious nodes only result in discarding of compromised paths and selection of new paths from the stored list, resulting in a slight drop of the communication delay. It is also observed that even with 25% of node compromise, the communication delay is as low as 30.4 seconds.

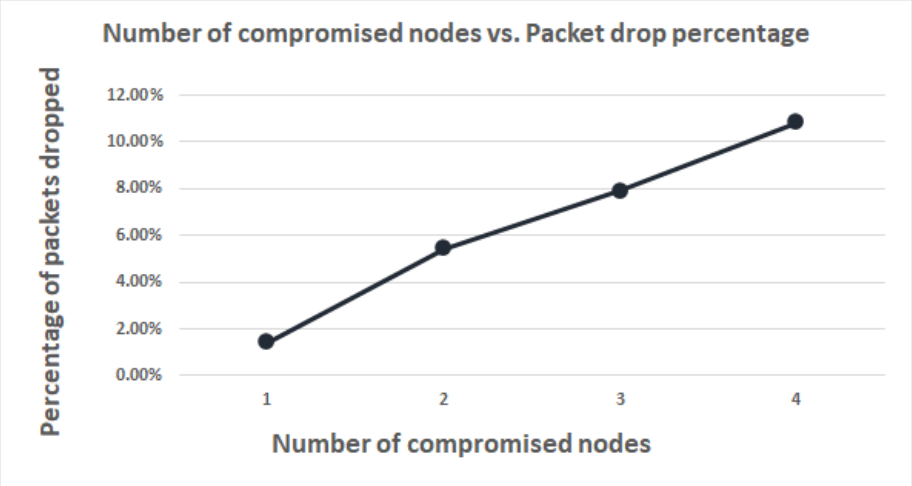

Fig. 5.7. Percentage of node compromise vs. Percentage of Packets Dropped

Fig. 5.7. shows the number of compromised nodes in the network vs. the percentage of packets dropped. It is observed that even when 40 EHRNs are compromised, the packet drop percentage is only 12.5%. Each gateway selects a portion of the total number of nodes in the network in the four most trusted paths to PDCs and as a result most of the other nodes

remain idle and do not take part in the communication at that time. If such idle nodes are compromised, the communication of PMU data to the CCs will not be harmed. Moreover, such nodes will not be selected in the trusted paths in later rounds. However, if an already selected node is compromised, it will result in packet drop in the next hop only. After the next hop of the compromised node, that false data is not forwarded, and also immediate action is taken to remove the malicious node from the network. Therefore, the packet drop is not very high in the proposed routing technique.

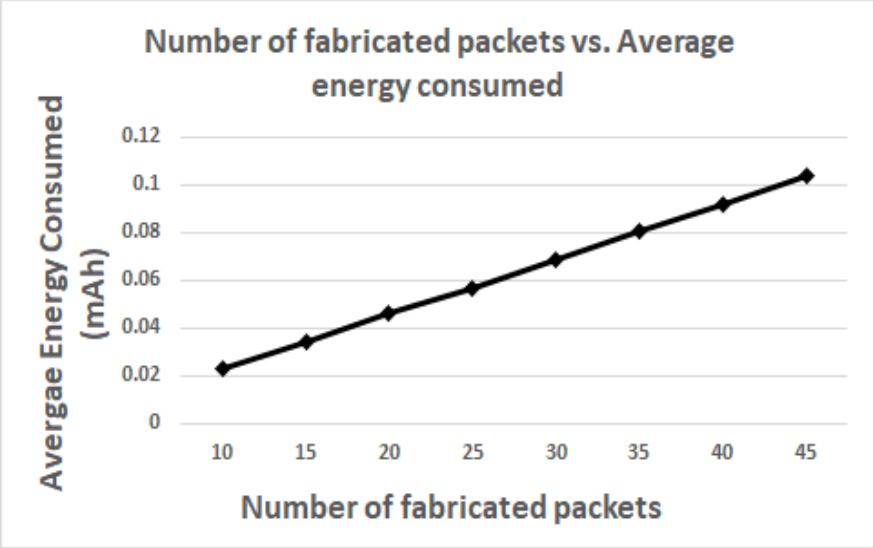

Fig. 5.8. Number of Fabricated Packets vs. Average Energy Consumed

In fig. 5.8., the average energy consumed by each active node vs. the number of fabricated packets is shown. The average energy consumed by each active EHRN when FDI attack takes place, increases as the number of fabricated packets increase in the network. However, the consumption of energy is very nominal by each active node for resending the data packets that are fabricated before. When there are 45 fabricated data packets in the network, the average energy consumed by each node is 0.104 mAh.

Chapter 6

OPTIMAL COST NETWORK DESIGN FOR BOUNDED DELAY DATA TRANSFER FROM PMU TO CONTROL CENTER USING HIGH BANDWIDTH CHANNELS

The Rooted Delay Constrained Minimum Spanning Tree (RDCMST) problem studied in [9] in a topological setting, is formalized in this chapter and studied in a geometric setting.

Although most of the prior works [36]-[38] have been focuses on topological setting, the study by Ho.et. al [39] is an exception. The primary focus of their study is finding a minimum diameter spanning tree (MDST) of a set of n points in the Euclidean plane. Given a set *P* of *n* points, $P = \{p_1,\ p_2, \ldots p_n\}$, the Euclidean graph *G* induced by *P* is a weighted complete graph $G = (V, E)$, where the weight of the edge $(p_i, p_j) \in E$ is the Euclidean distance between the points p_i and p_j. In [39] they formulate the BRBCST problem in the following way: Given a graph $G = (V, E)$, a cost (or weight) function $W(e) \in Z^+$ for all $e \in E$, a distinguished vertex $v \in V$ as the *root*, and positive integers *R* and *C*, find a spanning tree *T* for *G* such that $\sum_{e \in T} \leq C$ and $\sum_{e \in p} \leq R$ for all simple paths $p \in T$ starting from the root *v*. They prove that the BRBCST problem is NP-complete.

The BRBCST problem [39] was studied in a topological setting by Salama et.al in [36] and Xue in [38]. In [36] the problem was referred to as Delay-Constrained Minimum Spanning Tree (DCMST) and in [38] it was referred to as Delay Constrained Multicast Tree (DCMT). As the name Rooted Delay-Constrained Minimum Spanning Tree (RDCMST) seems to be the most appropriate way to describe the problem, it is referred to

as RDCMST in this chapter. The authors in [36] prove that the RDCMST problem NP-complete in a topological setting and present a heuristic based on the Prim's Spanning Tree algorithm [40] . As DCMT problem studied in [38] is very close to RDCMST problem here, an elaborate difference between the two is also presented in this chapter.

In this chapter, (i) the necessary and sufficient condition for the existence of a solution for the RDCMST problem is established, (ii) it is demonstrated that that both the heuristic algorithms may fail to find the optimal solution for some problem instances, (iii) some conditions are characterized on the input data which will ensure that the heuristic algorithms will find the optimal solution, (iv) it is also demonstrated that under some pathological condition, the ratio between each of the heuristic algorithms and the optimal solution can be arbitrarily large. An Integer Linear Programming formulation for the RDCMST problem is provided in this chapter for the computation of the optimal solution. Performance of the heuristic algorithm: M_Prim is evaluated with real substation location data of Arizona.

6.1. Problem Formulation

The input for the RDCMST problem is a set of points, $P = \{p_1, \dots p_n\}$ (locations of substations) are given on a two-dimensional plane (or a sphere), one of which is a distinguished point (say, p1), as it corresponds to the location of the control center, and an acceptable delay threshold value δ. A weighted, complete graph $G = (V, E)$ is constructed, where each node $v_i \in V$ corresponds to a point $p_i \in P$. Since the node v_i and the point p_i has a one-to-one correspondence, the terms node and point are used interchangeably in

this chapter. The weight $w(v_i, v_j)$ of an edge $(v_i, v_j) \in E$, is either the Euclidean distance (planar surface) or Spherical distance, measured by Haversine formula [41] (spherical surface e.g., the surface of the earth). It is assumed that $w(v_i, v_j)$ represents both cost and delay of the link (v_i, v_j).

It is to be noted that this is a simplification, as in a realistic communication system, the link length is only one of the factors that will determine the cost and the delay associated with that link. In case of link cost, another factor could be link type - wired/wireless. Similarly for delay, the other factors may be transmission rate, link bandwidth, queuing delay etc., whereas the link length can only determine the propagation delay. However, this simplified model is widely used in literature on communications networks as many efficient algorithms for more complicated models are based on efficient algorithms for this simplified model.

The objective of the RDCMST problem is to construct the least cost network subject to the constraint that the length of the path from any node $v_i, 2 \leq i \leq n$ to the node v_1 does not exceed the delay threshold δ. This actually translates to construction of the least cost spanning tree for the graph $G = (V, E)$ subject to the constraint that path length from all nodes $v_i, 2 \leq i \leq n$ to v_1 is at most δ. It is to be noted that the cost of the spanning tree is equal to the sum of the weights of all the edges that make up the spanning tree, and delay of a node v_i to the node v_1, is equal to the sum of the weights of all the edges that constitute the path from v_i to v_1 in the spanning tree.

6.2. Difference between DCMT and RDCMST problems

As the DCMT problem studied in [38] is similar to the RDCMST problem studied in this chapter, to avoid any confusion, the difference between the two is illustrated in this section with the help of an example. The DCMT problem is defined as follows:

Let $\alpha > 1$ be a given constant. A multicast tree T is called $\alpha - optimal$ if $d(s, v, T) \leq \alpha . d(s, v,\ SPT)$ is true for every node $v \in D$. The delay-constrained multicast tree problem asks for a given constant α. The notations used in [38] are as follows:

- $\pi(u, v, T)$: Unique path from *u* to *v* in *T*.
- $d(u, v, T)$: Delay on the path $\pi(u, v, T)$.
- s: Source (Root) node of the multicast tree *T*.
- D: Destination nodes of the multicast tree *T*.
- SPT: A multicast tree $T(s, D)$ such that $d(s, v, T)$ is the smallest possible delay for every node $v \in D$.

The RDCMST problem can be viewed as a DCMT problem by setting $s = v_1$ and $D = V - \{v_1\}$. The difference between the two is illustrated with the example shown in fig. 6.1., which shows four points with their (x, y) coordinates in a two-dimensional plane. The Euclidean distance between every pair of points is shown as the weight on the line (edge) connecting the pair of points (nodes). The point 1 is the special node (root of the spanning tree). As per Cayley's formula the number of spanning trees of a *n* node complete graph with labeled vertices is n^{n-2} [42].

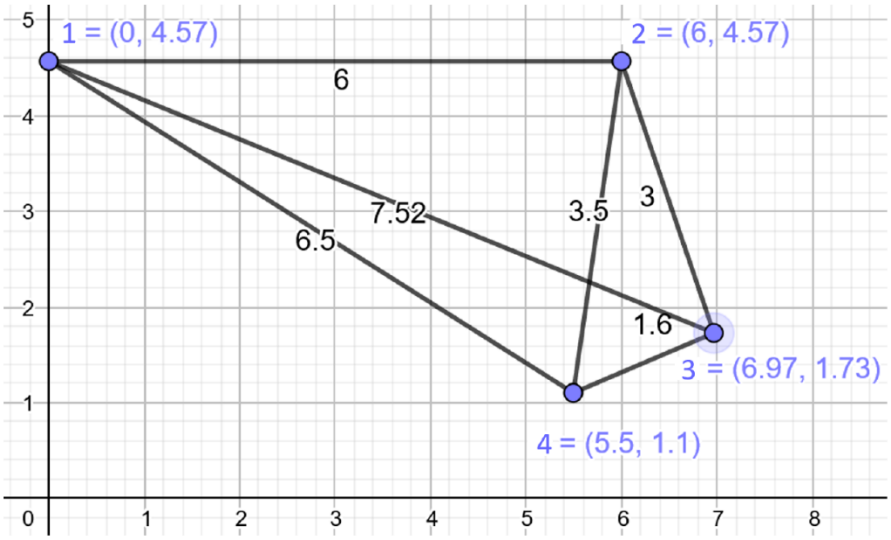

Fig. 6.1. RDCMST Problem Instance with 4 Points in a 2-Dimensional Plane

Trees	Edges	Delays	Cost
$T_1(+)$	(1, 2), (1, 3), (1, 4)	(6, 7.52, 6, 5)	20.02
$T_2(\#)$	(1, 2), (2, 3), (2, 4)	(6, 9, 9.5)	12.5
T_3	(1, 3), (2, 3), (2, 4)	(10.52, 7.52, 14.02)	14.02
T_4	(1, 4), (2, 3), (2, 4)	(10, 13, 6.5)	13
T_5	(1, 2), (2, 3), (3, 4)	(6, 9, 10.6)	10.6
T_6	(1, 3), (2, 3), (3, 4)	(10.52, 7.52, 9.12)	12.02
T_7	(1, 4), (2, 3), (3, 4)	(11.1, 8.1, 6.5)	11.1
T_8	(1, 2), (2, 4), (3, 4)	(6, 11.1, 9.5)	11.1
T_9	(1, 3), (2, 4), (3, 4)	(12.62, 7.52, 9.12)	12.62
T_{10}	(1, 4), (2, 4), (3, 4)	(10, 8.1, 6.5)	11.6
$T_{11}(+)$	(1, 2), (1, 4), (3, 4)	(6, 8.1, 6.5)	14.1
T_{12}	(1, 2), (1, 3), (3, 4)	(6, 7.52, 9.12)	15.12
$T_{13}(+)$	(1, 2), (1, 4), (2, 3)	(6, 9, 6.5)	15.5
T_{14}	(1, 3), (1, 4), (2, 3)	(10.52, 7.52, 6.5)	17.02
T_{15}	(1, 2), (1, 3), (2, 4)	(6, 7.52, 9.5)	17.02
T_{16}	(1, 3), (1, 4), (2, 4)	(10, 7.52, 6.5)	17.52

Table 6.1. Spanning Trees Corresponding to 4 Points in Fig. 6.1.; Column 3 Triple (X, Y, Z) Indicates the Data Transfer Delays from Points 2, 3 and 4 to Point 1 respectively

In this example with *n=4*, the number of spanning trees is $4^{4-2} = 16$. These 16 trees are listed as T_1 through T_{16} in Table 6.1. As *n=4*, there will be $n - 1 = 3$ edges in each one of the spanning trees and these edges are listed in column 2 of Table 6.1. The delays encountered by data to arrive to node 1 from the nodes 2, 3, and 4 are listed in

column 3 of Table 6.1. It may be recalled here that a delay from a node $v \in \{2, 3, 4\}$ is measured by the length of the path from node v to node 1. The length of the path from $2 \rightarrow 1$, $3 \rightarrow 1$ and $4 \rightarrow 1$ in the spanning tree T_1 (which is made up of edges $(1, 2), (1,3)\ and\ (1,4)$) is 6, 7.52 and 6.5 respectively, and is shown in column 3 of Table 6.1. Similarly, the path lengths from $2 \rightarrow 1$, $3 \rightarrow 1$ and $4 \rightarrow 1$ in the spanning tree T_2 (which is made up of edges $(1, 2), (2,3)\ and\ (2,4)$) is 6, 9 and 9.5 respectively. The cost of a spanning tree is measured by the sum of all the edge weights constituting the tree. Accordingly, the cost of T_1 and T_2 are 20.02 and 12.5 respectively and is shown in column 4 of Table 6.1. The edges, delays and the costs associated with the other 14 trees, T_3 to T_{16} are also shown in the same table.

It may be recalled that the objective of DCMT problem is to find, for a given input parameter α, an $\alpha - optimal$ multi-cast tree, where a multicast tree T is defined to be $\alpha - optimal$ if $d(s, v, T) \leq \alpha . d(s, v, SPT)$ for every node $v \in D$. It may be verified that in the example shown in fig. 6.1., the tree (T_1) with the network design cost = 20.02 is the SPT, as $d(1, v, T_1)$ is the smallest possible for every node $v \in \{2, 3, 4\}$. The delays from $2 \rightarrow 1, 3 \rightarrow 1$ and $4 \rightarrow 1$ in the tree T_1 are 6, 7.52 and 6.5 respectively. If the input parameter α is given as 1.3, the DCMT problem objective becomes finding the least cost spanning tree, where the delay bound for the nodes 2, 3 and 4 to reach node 1, will be $6 \times 1.3 = 7.8,\ 7.52 \times 1.3 = 9.776$ and $6.5 \times 1.3 = 8.45$ respectively. It may be seen from Table 6.1, that only three trees, T_1, T_{11} and T_{13} (marked (+)), can satisfy these delay requirements and T_{11}being the least expensive among these three, the DCMT problem will choose T_{11} as the solution, with a network design cost of 14.1. While DCMT takes in α as

an input parameter, the RDCMST takes in a different parameter α sets a limit on the amount of deviation the DCMT is allowed to make from the optimal delay, whereas α sets a limit on the amount of absolute delay that is acceptable (i.e. respective of the value of the optimal delay). In order to make a meaningful comparison between DCMST and RDCMST, when $\alpha = 1.3$, δ is set to be equal to largest of the three delay constraints in DCMT problem instance (i.e., 7.8, 9.776 and 8.45). If $\delta = 9.776$, then the solution of the RDCMST problem will be the tree T_2 (marked #) with a network design cost of 12.5. This example clearly demonstrates that an algorithm designated to solve the DCMT problem may not solve the RDCMST problem.

Lemma 1: The necessary and sufficient condition for existence of a solution for the RDCMST problem is $max_{2\leq i\leq n} dist(p_1, p_i) \leq \delta$, where $dist(p_1, p_i)$ represents the Euclidean (Spherical) distance between the points p_1 and p_i.

Proof: The RDCMST problem instance is specified by a set of points $P = \{p_1, \dots p_n\}$ and a delay threshold δ.

Necessary: Suppose that for a given problem instance of the RDCMST problem, $D = max_{2\leq i\leq n} dist(p_1, p_i)$, and the point furthest from the point p_1 in P is p_j. If the delay threshold for the problem instance, $\delta < D$, then there is no way the data from p_j can arrive at p_1 without violating the delay threshold δ. Thus, in order to have a solution to an RDCMST problem, the condition $\delta \geq D$ must be satisfied.

Sufficient: If $\delta \geq D$, then if the point p_1 is directly connected to every other point $p_j \in P$, $2 \leq j \leq n$, then such a network will satisfy the delay threshold constraint. Thus, a solution to the RDCMST problem must exist when $\delta \geq D$.

6.3. RDCMST Integer Linear Program

Data has to flow from the PMUs at the substations (nodes v_2 through v_n) to the control center (node v_1) within the specified delay threshold δ. There are $n-1$ flows from the nodes $v_i, 2 \leq i \leq n$ to node v_1. These $n-1$ flows are denoted as $f^2, \ldots, f^n \; and$ the directed edge from the node v_i to the node v_j as $e_{i,j}$. The cost and the delay associated with the edge $e_{i,j}$ are denoted as $C_{i,j}$ and $D_{i,j}$ respectively. The nodes adjacent to the node $v_i, 1 \leq i \leq n$ is denoted by $N(v_i)$. The binary variable $f_{i,j}^k$ is used to defined as follows:

$$f_{i,j}^k = \begin{cases} 1, \text{if the flow } f^k \text{uses the directed edge } e_{i,j} \\ 0, otherwise \end{cases} \tag{6.1}$$

Total flow on the edge $e_{i,j}$ is denoted by $F_{i,j}$ and is given as the sum of all the flows on that edge,

$$F_{i,j} = \sum_{k=2}^{n} f_{i,j}^k \tag{6.2}$$

A binary variable $X_{i,j}$ is used, such that:

$$X_{i,j} = \begin{cases} 1, & if\ F_{i,j} \geq 1 \\ 0, otherwise \end{cases} \tag{6.3}$$

The objective of the RDCMST problem, which is to minimize the network design cost, can be expressed with the following objective function:

Objective Function: $Minimize\ \sum_{i=1}^{n}\sum_{j=1}^{n} C_{i,j}X_{i,j}$ (6.4)

i. The Delay Constraint is denoted as: $\forall\ k, 2 \le k \le n, \sum_{i=1}^{n}\sum_{j=1}^{n} D_{i,j}f_{i,j}^{k} \le \delta$ (6.5)

ii. Constraint at node v_1 for the flow f^k from v_k to v_1:

$$\forall\ k, 2 \le k \le n, \sum_{j\in N(v_1)} f_{j,1}^{k} - \sum_{j\in N(v_1)} f_{1,j}^{k} = 1 \quad (6.6)$$

iii. Constraint at node v_k for the flow f^k from v_k to v_1:

$$\forall\ k, 2 \le k \le n, \sum_{j\in N(v_1)} f_{j,k}^{k} - \sum_{j\in N(v_1)} f_{k,j}^{k} = -1 \quad (6.7)$$

iv. Constraints at node v_l, $\left(v_l \in \{V - \{1, k\}\}\right)$ for the flow f^k from $v_k \longrightarrow v_1$:

$$\forall\ k, 2 \le k \le n, \sum_{j\in N(v_l)} f_{j,l}^{k} - \sum_{j\in N(v_l)} f_{l,j}^{k} = 0 \quad (6.8)$$

v. Constrains to relate $F_{i,j}$ with $X_{i,j}$:

$$F_{i,j} \le (n-1)X_{i,j} \quad (6.9)$$

6.4. Modified Prim (M_Prim) Algorithm

In this section, first, the modified version of the well-known Prim's Algorithm [40] for computation of a minimum spanning tree of a graph, called M_Prim is presented. Then, (i) it is shown that M_Prim algorithm for some problem instances may fail to find the optimal solution, (ii) the conditions are categorized under which M_Prim is guaranteed to find the optimal solution, and (iii) finally it is demonstrated that under some pathological condition, the ratio between the solution produced by M_Prim and the optimal solution can be arbitrarily large.

Algorithm 6.1: M_Prim(P, δ)

Result: M_Prim_RDCMST(P, δ)

1. Construct a complete graph $G = (V, E)$ where each node v_i corresponds to a point $p_i \in P$ and weight $w(v_i, v_j)$ on the edge $(v_i, v_j) \in E$ is the Euclidean distance between the points p_i and p_j.
2. Partition the vertex set V into subsets V_1 and V_2 and initialize them as $V_1 \leftarrow \{v_1\}$ and $V_2 \leftarrow V - \{v_1\}$.
3. Initialize an edge set E_1 as an empty set, $E_1 = \emptyset$,
4. while $V_2 \neq \emptyset$ do,

begin

(i) Find the smallest weight edge $e \in E, e = (v_i, v_j)$ such that $v_i \in V_1, v_j \in V_2$ and in the graph $G_1 = (V_1, E_1 \cup \{(v_i, v_j)\})$ the path length from v_1 to v_j is at most δ. If no such node v_j exists, then there is no solution to the input RDCMST problem instance (P, δ).

(ii) $V_1 = V_1 \cup \{v_j\}$

(iii) $V_2 = V_2 - \{v_j\}$

(iv) $E_1 = E_1 \cup \{(v_i, v_j)\}$

(v) $E = E - \{(v_i, v_j)\}$

end

Output: The graph $G_1 = (V_1, E_1)$, which is the RDCMST formed using M_Prim

Claim 1: The M_Prim Algorithm may fail to find an optimal solution for some RDCMST problem instances.

Proof: For instance, in fig. 6.1., if the delay threshold $\delta = 10$, then M_Prim will select the edge (1, 2) first and (2, 3) second. It will then try to select the edge (3, 4) (as it has the smallest weight), but will *reject* it, as in this case the delay from the node 4 to node 1 will be 1.6 + 3 + 6 = 10.6, exceeding the delay threshold. After rejecting the edge (3, 4), it will examine the edge (2, 4) (as it's the next smallest weighted edge) and will *accept* it, as in this case, the delay from node 4 to node 1 is 3.5 + 6 =9.5, which is within the delay

threshold. Thus, the cost of the network designed by M_Prim will be 6 + 3 + 3.5 = 12.5. However, a network constructed by choosing the edges (1, 4), (2, 4) and (3, 4), would have satisfied the delay constraint and would have been of lower cost, 6.5 + 3.5 + 1.6 =11.6. This example demonstrates that for some instances, M_Prim may fail to find the optimal solution.

In the following, some conditions are characterized under which M_Prim is guaranteed to produce the optimal solution. Suppose D is an $n \times n$ matrix, where the i^{th} row, j^{th} column entry $(d_{i,j})$ represents the Euclidean distance between points p_i and p_j. Suppose the following terms are defined as:

$$d_{smallest} = Smallest\ number\ in\ D = min_{1 \le i,j \le n; i \ne j} d_{i,j}$$

$$d_{2nd-smallest} = 2nd\ smallest\ number\ in\ D$$

$$Max_{Dist(i)} = Distance\ of\ the\ farthest\ point\ p_j\ from\ the\ point\ p_i = max_{1 \le j \le n} d_{i,j}$$

$$\Delta_{small} = d_{smallest} + d_{2nd-smallest}$$

$$\Delta_{large} = \sum_{i=1}^{n} Max\_Dist\ (i)$$

Claim 2: The M_Prim Algorithm is guaranteed to produce the optimal solution to the RDCMST problem if the delay threshold is $\delta < \Delta_{small}$ or $\delta \geq \Delta_{large}$.

Proof: The Lemma 1 in section 6.2 established that $\delta \geq Max\_Dist(1)$ is the necessary and sufficient condition for existence of a solution for the RDCMST problem. If $Max\_Dist\ (1) \leq \delta \leq \Delta_{small}$, then the solution to RDCMST can be constructed by

connecting every point p_i, $2 \leq i \leq p_n$ directly to the point p_1, as otherwise if the path from p_k, then the length of this path would exceed δ. Accordingly, when $\delta \leq \Delta_{small}$, both in the optimal as well as M_Prim solution, each point p_i must be directly connected to p_1, and the two solutions will be identical. An example of the problem instance where $Max\_Dist(1) \leq \delta \leq \Delta_{small}$: Consider three points $\{p_1, p_2, p_3\}$ with (x, y) co-ordinates (0, 4), (0, 0) and (3, 0) respectively and $\delta = 6$. In this case, $Max_{Dist(1)} = 5 \leq \delta = 6 \leq \Delta_{small} = 7$.

As M_Prim solution will select exactly $n - 1$ edges from the edge set E of the complete graph $G = (V, E)$, the complete graph created from the set of points P, if $\delta \geq \Delta_{large}$, no matter which $n - 1$ edges are selected by the M_Prim algorithm, the delay threshold will never be exceeded. In this situation the output of the M_Prim algorithm will be identical to the output of the Prim's algorithm, which is guaranteed to produce the optimal spanning tree. As the minimum spanning tree (MST) (i.e., output of the Prim algorithm) is a lower bound of both the optimal solution of the RDCMST problem (Opt_RDCMST) and M_Prim solution (M_Prim_RDCMST); and (Opt_RDCMST) is a lower bound of (M_Prim_RDCMST), the relationship between MST, Opt_RDCMST and M_Prim RDCMST, is $MST \leq Opt\_RDCMST \leq M\_Prim\_RDCMST$. If $\delta \geq \Delta_{large}$, then $MST = M\_Prim\_RDCMST$, which in turn also implies that $Opt\_RDCMST = M\_Prim\_RDCMST$.

Claim 3: The ratio between the solution produced by the M_Prim algorithm and the Optimum Solution of the RDCMST problem can be arbitrarily large.

Proof: This claim can be substantiated with the help of an example. Consider a set of points $P = \{p_1, \ldots p_n\}$ on a two-dimensional plane with their (x, y) co-ordinates. In fig. 6.2, $n = 11$ and their (x, y) co-ordinates are shown next to the points. In general, the number of points in such a set of points is $n = 2k + 1$, where k is an integer. The co-ordinates of the points are (i, i) and $(i, i + 1)$ for $0 \leq i \leq k$. If point p_1 with (x, y) co-ordinates (0, 0) is taken as the distinguished point (i.e., the location of the CC), the point p_n is the furthest from p_1. If the delay threshold δ is set equal to the distance between p_1 and p_n, then $\delta = k \times \sqrt{2} = 1.41k$. The optimal solution to this RDCMST problem instance is shown in Fig. 6.2. Thus, the cost of the optimal solution is $k \times 1 + k \times \sqrt{2} = (\sqrt{2} + 1)k$, $i.e. Opt_{RDCMST} = (\sqrt{2} + 1)k$.

It may be noted that M_Prim and Optimal solution selects a set of $n - 1$ edges from the complete graph created from the point set P. Thus, the ratio $R = \frac{\sum_{i=1}^{n-1} w(e_i')}{\sum_{i=1}^{n-2} w(e_i)}$, where $w(e_i')$ and $w(e_i)$, represents the weights of the edges selected by the Optimal and M_Prim algorithm respectively. As the M_Prim algorithm always picks the least cost link, as long as its selection doesn't violate the delay threshold δ, it will keep on selecting edges with cost 1 like (1, 2), (2, 3),…, (fig. 6.3.), till the number of such edges exceeds $\lfloor k\sqrt{2} \rfloor$. Of the $n - 1$ edges selected by the M_Prim, $\lfloor k\sqrt{2} \rfloor$ edges will be of cost 1, with a total cost of $\lfloor k\sqrt{2} \rfloor$. The edges with cost 1 are referred to as *short edges* in this chapter and those with cost greater than 1 as *long edges* (LE). The remaining $t = (n - 1) - \lfloor k\sqrt{2} \rfloor = 2k - \lfloor k\sqrt{2} \rfloor$ edges selected by M_Prim will be LE.

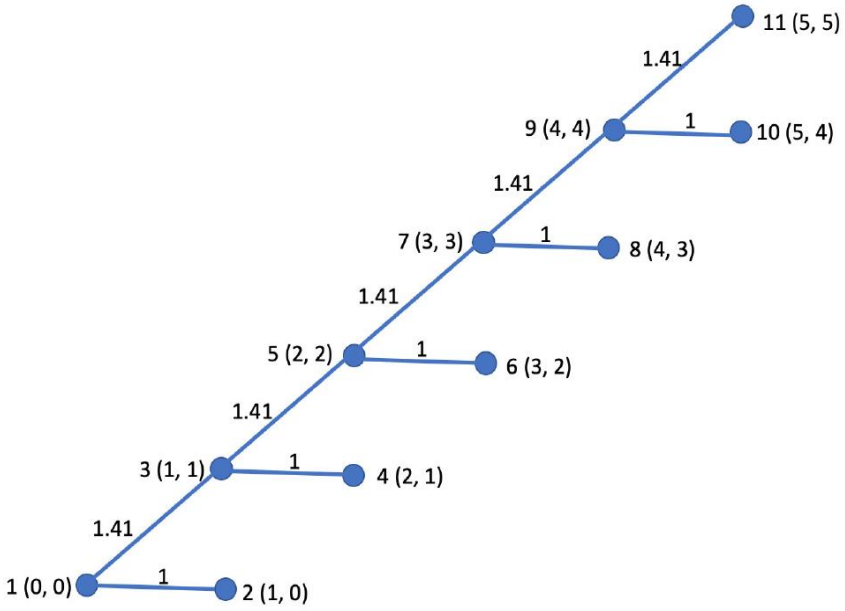

Fig. 6.2. Opt_RDCMST Solution with 11 Points

As shown in fig. 6.4., the odd and even numbered points may be thought of belonging to the Upper and Lower Diagonal (UD and LD) respectively. A set of four points is defined with co-ordinates $\{(i,i),(i+1,i),(i+1,i+1),(i+2,i+1)\}$ to be a Block. The LE can be of two types− (i) that are connecting two points on the same (Upper or Lower) diagonal and, (ii) connecting two points on different diagonals. These two types are referred to as Same Side Long Edges (SSLE) and Different Side Long Edges (DSLE). The length (cost) of a long edge is dependent on the number of Blocks it spans. If a Long Edge $LE_{i,}1 \leq i \leq t$ spans s_i Blocks, its length is denoted by $LE(s_i)$. If a Long Edge $LE_{i,}$ is of the type SSLE and it spans s_i Blocks, then $LE(s_i) = s_i\sqrt{2}$. This can be verified in fig. 6.3, where the SSLE type LE connecting nodes 9 and 5 has length $2\sqrt{2}$ (spans 2 Blocks) and, the LE connecting the nodes 11 and 1 has length $5\sqrt{2}$ as it spans 5 Blocks.

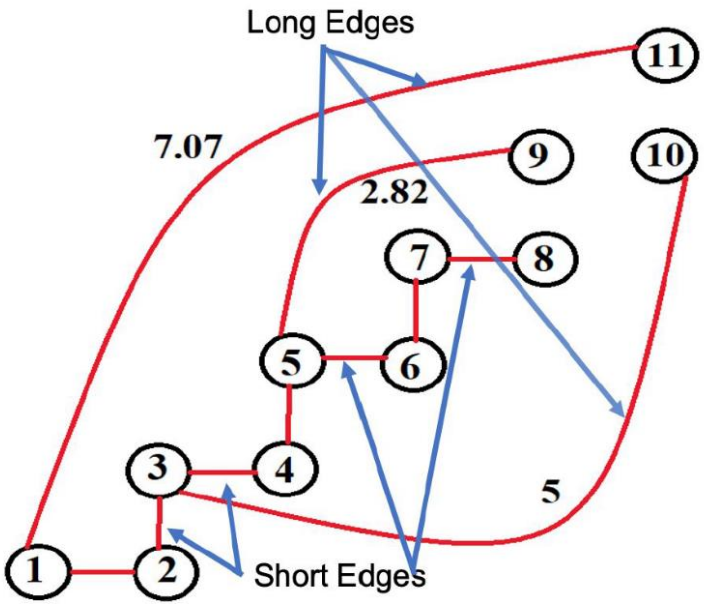

Fig. 6.3. M_Prim_RDCMST Solution with 11 Points

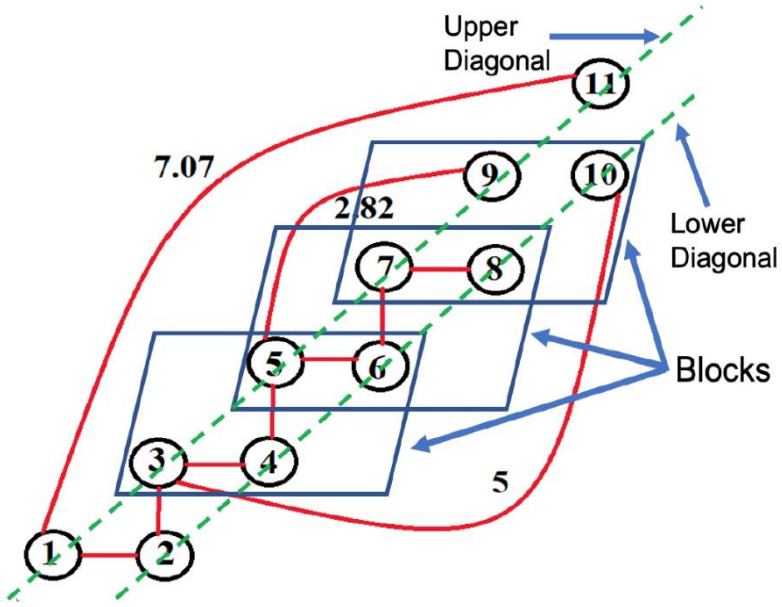

Fig. 6.4. Blocks and Diagonals in the RDCMST Problem Instance

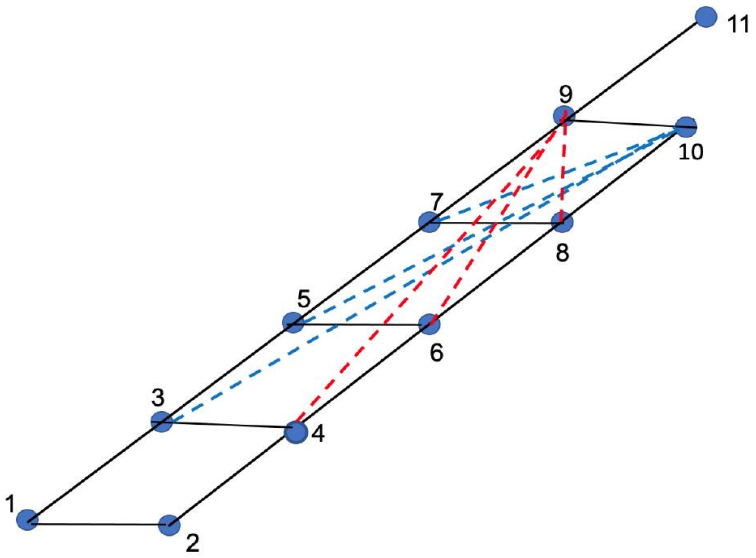

Fig. 6.5. Distance between Points on the Upper and Lower Diagonals

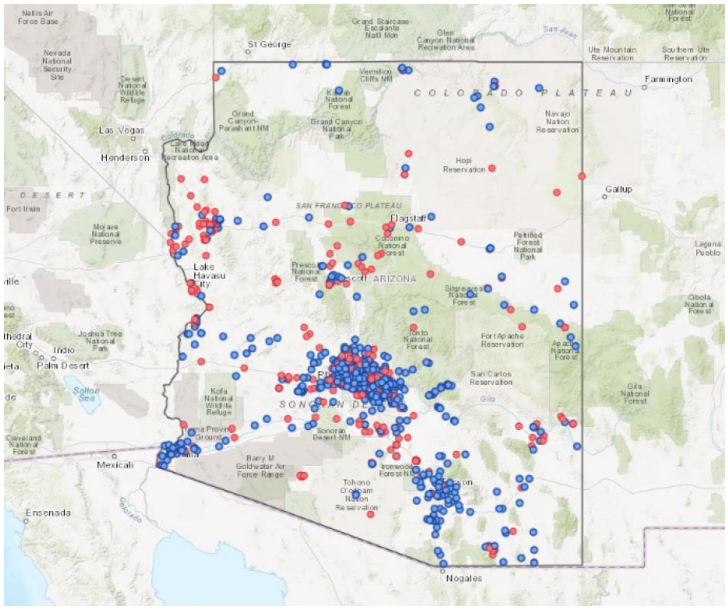

Fig. 6.6. Distribution of Points in Arizona

If a Long Edge LE_i is of the type DSLE and spans s_i Blocks, then its length depends not only on s_i, but also on the locations of the nodes it's connecting. When it's connecting two nodes *u* and *v*, the higher indexed node is considered as the source and the lower indexed node as the destination. If LE_i is spanning s_i Blocks, $LE(s_i)$ may be different, depending

on whether the source is on UD and the destination is on LD, or the other way around. These two types of DSLEs are referred as Types I and II respectively. It can be verified in fig. 6.2, where the length of the Type I DSLE connecting node 9 and 6 is different from length of the Type II DSLE connecting the nodes 10 and 5, although both the Les are spanning same 2 Blocks. It can be shown through simple geometric calculation for the diagonal of the parallelogram comprising of s_i Blocks (fig. 6.5), that $LE(s_i)$ for Types I and II DSLEs are $\sqrt{2s_i^2 - 2s_i + 1}$ and $\sqrt{2s_i^2 + 2s_i + 1}$ respectively. This can be verified in fig. 6.3, where the Type II DSLE connecting the nodes 10 and 3, spanning 3 Blocks has length $LE(3) = \sqrt{2s_i^2 + 2s_i + 1} = \sqrt{2 \times 3^2 + 2 \times 3 + 1} = 5.$

Thus, the ratio *R* between the M_Prim and the Optimal is,

$$R = \frac{M\_Prim\_RDCMST}{Opt\_RDCMST} = \frac{w(e_1')+\cdots+w(e_{n-1}')}{w(e_1)+\cdots+w(e_{n-1})} = \frac{\lfloor k\sqrt{2} \rfloor + \sum_{i=1}^{t=(2k-\lfloor k\sqrt{2} \rfloor)} LE_i}{k+k\sqrt{2}} = \frac{\lfloor k\sqrt{2} \rfloor + (LE(s_1)+\cdots+LE(s_{t-1})+LE(s_t))}{k(1+\sqrt{2})} \quad (6.10)$$

As noted earlier, M_Prim will select $t = 2k - \lfloor k\sqrt{2} \rfloor$ LEs connect node v_1 to nodes v_{n-t+1} through v_n. In fig. 6.3., n=11, k=5, t=3, and three LEs, LE_1, LE_2, LE_3, connecting node 1 to node 9, 10, 11 respectively. Each LE_i spans s_i Block. In fig. 6.3, $s_1 = 2, s_2 = 3, s_3 = 5$. If LE_i and LE_{i+1} connects node *u* and *v* to v_1 respectively, then *v* is further away than *u* from v_1. This implies LE_{i+1} spans more Blocks than LE_i. In other words, $s_i < s_{i+1}$. More generally, $s_1 < s_2 < \cdots < s_{t-1} < s_t$. Moreover, $s_1 > 0$ and $s_t = k$. As noted earlier, if LE_i is of the type SSLE, then $LE(s_i) = s_i\sqrt{2}$.

If it's of the type DSLE, then $LE(s_i) = \sqrt{2s_i^2 + 2s_i + 1}$ or $\sqrt{2s_i^2 - 2s_i + 1} = \sqrt{2(s_i - 1)^2 + 2(s_i - 1) + 1}$. In both cases $LE(s_i) > (s_i - 1)\sqrt{2}$. As all s_i must be integers and $s_i < s_{i+1}, 1 \leq i \leq (t-1)$, even if s_1 through s_{t-1} are taken to be integers 1 through $t-1, \sum_{i=1}^{t} LE(s_i)$ will be $\Omega(t^2)$, which in turn is $\Omega(k^2)$. Thus, from equation (6.10) it follows:

$$R \geq \frac{\lfloor k\sqrt{2} \rfloor + (s_1\sqrt{2} + \cdots + s_{t-1}\sqrt{2} + s_t\sqrt{2})}{k(1+\sqrt{2})} \geq \frac{\lfloor k\sqrt{2} \rfloor + (s_1 + \cdots + s_{t-1} + s_t)\sqrt{2}}{k(1+\sqrt{2})} \geq \frac{\lfloor k\sqrt{2} \rfloor + \Omega(k^2)}{k(1+\sqrt{2})} =$$

$$\Omega(k) = \Omega(n) \tag{6.11}$$

This established the ratio between the solutions produced by the M_Prim algorithm and the RDCMST problem increases with n, and as such can be arbitrarily large.

6.5. Evaluation of the Modified Prim (M_Prim) Algorithm

The results of our experimental evaluation of the M_Prim algorithm with substation location data of Arizona is presented in this section. The latitude-longitude locations of substations in Arizona can be found from the U.S. Dept. of Homeland Security website [43] and are illustrated in fig. 6.6. The blue and red dots indicate the locations of operational and non-operational substations respectively. The total number of substations in Arizona is 892, of which 653 are operational. As PMUs are expensive, not all substations have it. It has been reported in [44] that PMUs installed in $20\% - 30\%$ of the substations are sufficient for full observability. Our experiments are conducted with substation locations in Phoenix and Tucson, two largest cities in Arizona. The number of substations in Phoenix and Tucson are 132 and 25 respectively.

Phoenix	Ratio = MRA_RDCMST(P, δ) /M_Prim_RDCMST(P, δ))					
Data Set	$\delta = 40$	$\delta = 50$	$\delta = 60$	$\delta = 70$	$\delta = 80$	$\delta = 90$
$DS_1(40\%)$	1.06	1.02	1.02	1.003	1	1
$DS_2(33\%)$	1.06	1.05	1.01	1	1	1
$DS_3(30\%)$	1.05	1.06	1.02	1.01	1	1
$DS_4(25\%)$	1.03	1.02	1.03	1.01	1	1
$DS_5(20\%)$	1.04	1.05	1.02	1.002	1	1

Table 6.2. Ratio between the M_Prim and Optimal Solutions for Phoenix

Tucson	Ratio = MRA_RDCMST(P, δ) /M_Prim_RDCMST(P, δ))					
Data Set	$\delta = 40$	$\delta = 50$	$\delta = 60$	$\delta = 70$	$\delta = 80$	$\delta = 90$
$DS_1(50\%)$	1.02	1	1.03	1.02	1	1
$DS_2(45\%)$	1.02	1.02	1.01	1	1	1
$DS_3(40\%)$	1.01	1.03	1.03	1.01	1	1
$DS_4(30\%)$	1	1	1.003	1	1	1
$DS_5(20\%)$	1	1	1	1	1	1

Table 6.3. Ratio between the M_Prim and Optimal Solutions for Tucson

The results of the experiments are presented in Table 6.2. and Table 6.3. The data sets DS_1 through DS_5 correspond to different percentage of the substations with PMUs. In Table 6.2., 40% next to DS_1 indicates that in data set 40% of 132 substations (53) are assumed to have PMUs. The delay threshold value δ was varied from 40 to 90. 53 substation locations from 132 were selected randomly thirty times. These thirty data sets may be viewed as $DS_{1,1}, DS_{1,2}, \ldots DS_{1,30}$. The ratios between the costs of the M_Prim and the Optimal solutions were computed for these thirty data sets for a specific δ value, and its average is presented in Tables 6.2. and Table 6.3. This value for Phoenix data $DS_1(40\%)$ and $\delta = 40$ is 1.06. The other entries in tables 6.2 and 6.3 were similarly computed. As the number of substations in Tucson were fewer, the percentage is varied from 20% to 50%, instead of 20% to 40% as was done for Phoenix. It can be inferred from these tables that M_Prim produces solutions with a very small amount of error in real data

sets (with the maximum deviation being 10%) and as such can be utilized for smart grid communication infrastructure design.

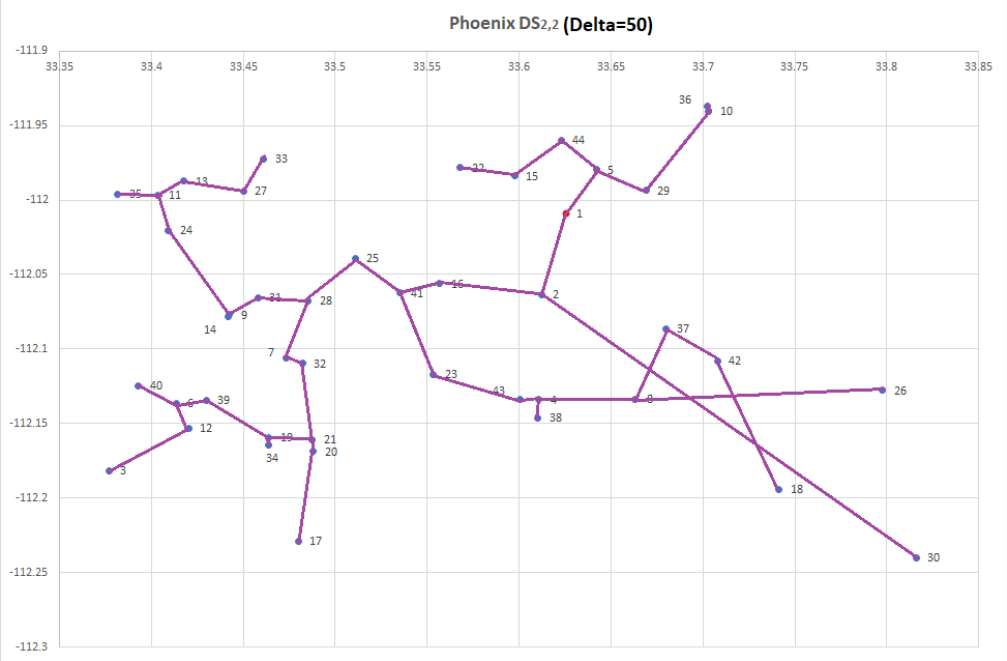

Fig. 6.7. M_Prim_RDCMST on Phoenix $DS_{2,2}$ with $\delta = 50$

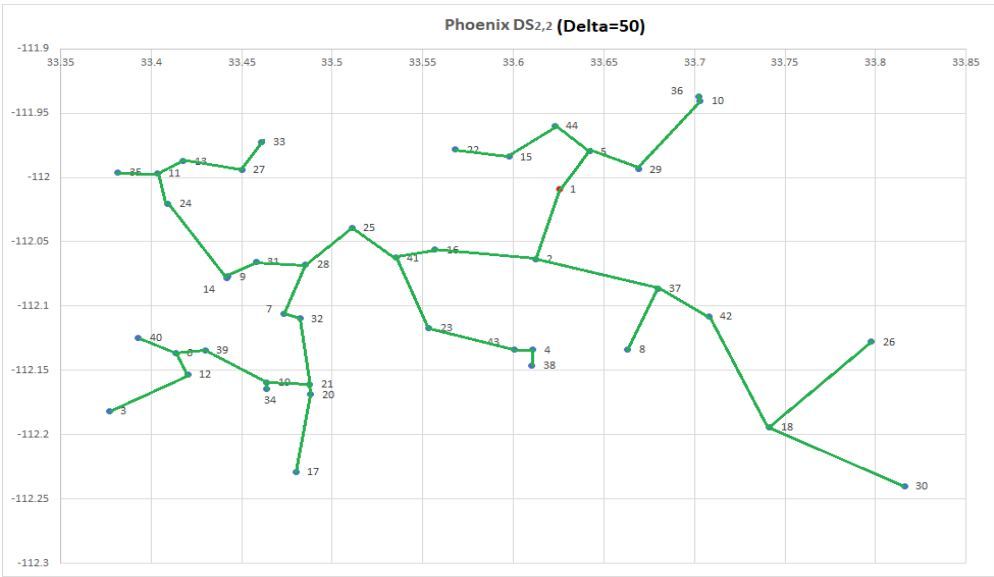

Fig. 6.8. Opt_RDCMST on Phoenix $DS_{2,2}$ with $\delta = 50$

Fig.6.7. shows the final RDCMST obtained using M_Prim algorithm considering one of the data sets of Phoenix with a total of 44 PMU containing substations denoted with blue points in the figure. The CC is point number 1, marked in red color. The delay value δ considered here is 50. The total cost of the tree created by the M_Prim algorithm for this data set is 182.45. Fig. 6.8. shows the RDCMST obtained using the ILP based solution for the same data set with same delta value and the total cost in this case is 159.61. Quite a big difference in the total ICT cost is noticed in this case. However, for the same data set, with $\delta = 80$, the RDCMST obtained is same for both the algorithms and it is shown in Fig 6.9. This also establishes that the performance of the M_Prim algorithm depends on the value of the delay constraint δ in some cases. It is to be noted that the RDCMST obtained for this data set with $\delta \geq 58.35$ using both the algorithms is same as MST. So, if the value of $\delta \geq MST\_Dist_{max}$, where MST_Dist_{max} is the maximum distance of a point in the graph from the root v_1 in the MST of that graph, then $RDCMST = MST$, no matter if M_Prim or the ILP based algorithm is used.

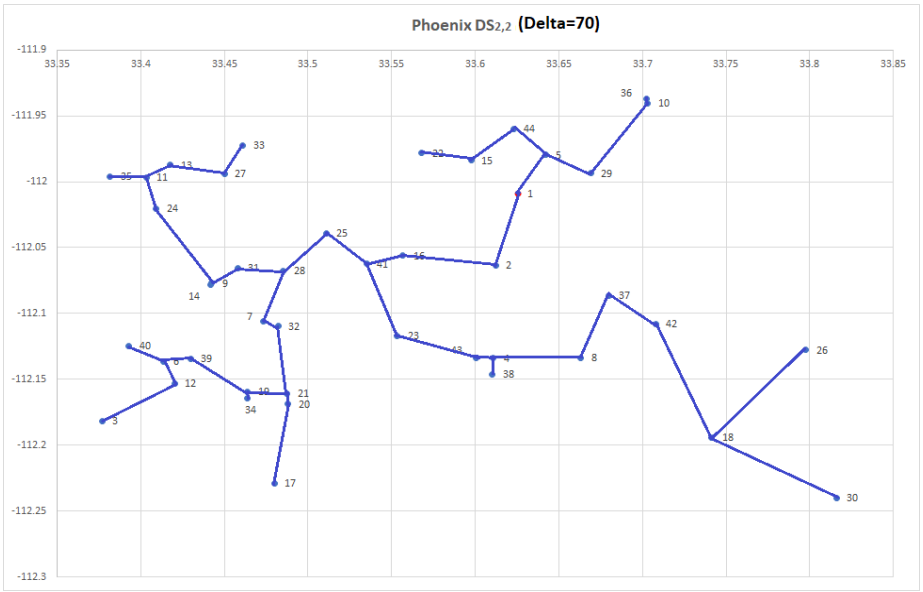

Fig. 6.9. Opt_RDCMST or M_Prim RDCMST on Phoenix $DS_{2,2}$ with $\delta \geq 58.35$

Chapter 7

DELAY CONSTRAINED COMMUNICATION NETWORK DESIGN FOR PMU TO MULTIPLE CONTROL CENTER DATA TRANSFER

Communication network plays an important role in a smart grid environment, as it must deliver data from the Phasor Measurement Units (PMUs) in the Substations (SSs) to the Control Center(s) (CCs) in real time. The PMUs generate data at a high rate that must arrive at the CC within an acceptable delay threshold. Accordingly, communication network design has received considerable attention from the researchers in recent times [9]. In [45] , this problem is considered where all the substations were sending data to a single CC and formalized it as Rooted Delay Constrained Minimum Spanning Tree (RDCMST) problem. As the number of substations in a geographic area is often large (e.g., Arizona has nearly nine hundred substations, fig. 6.6.), PMU data from the substations do not directly go to the Control Center (CC) and instead goes to multiple Local Controls Centers (LCC) within the specified delay threshold. The processed data from the LCCs is then sent to the CC. In this chapter, the results presented in Chapter 6 [45] is extended by considering Multiple Local Control Centers (MLCCs) where data from every PMU in the SSs must arrive at one of the multiple LCCs within the specified delay threshold. This setting gives rise to a new problem, where a Delay Constrained Spanning Forest needs to be created instead of a Delay Constrained Spanning Tree. The notion is formalized with the introduction of the Multi-Rooted Delay Constrained Minimum Spanning Forest (MRDCMSF) problem. In this chapter, (i) an optimal solution for the problem using Integer

Linear Programming, (ii) a Lagrangian Relaxation based solution and (iii) a heuristic solution with an innovative contention resolution mechanism is presented. Finally, the performance of the heuristic algorithm and the Lagrangian relaxation based solution is evaluated with real substation location data of Arizona.

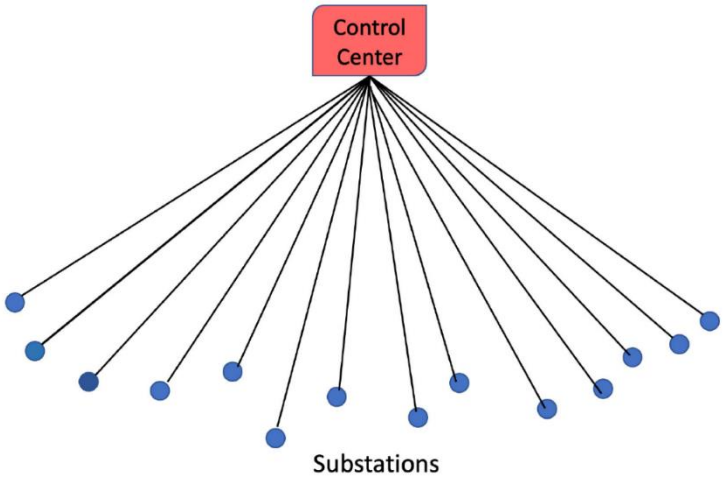

Fig. 7.1. SS to CC Direct Connections

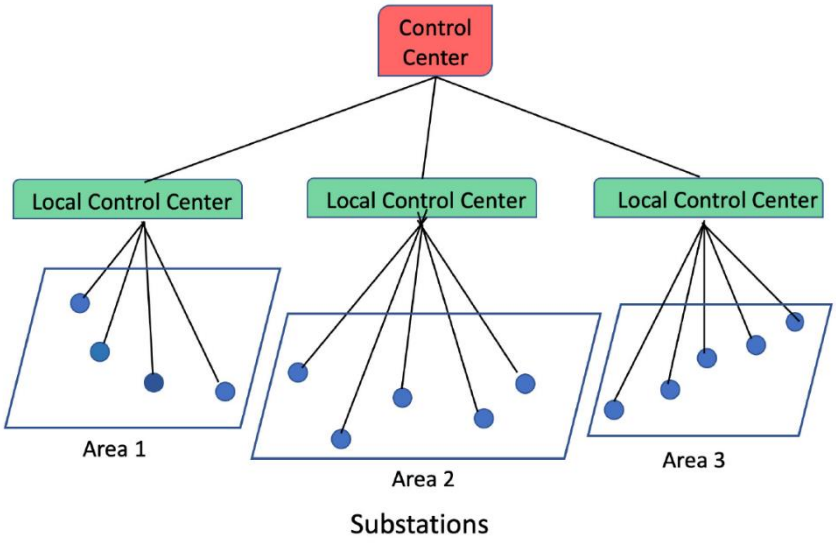

Fig. 7.2. SS to CC Through LCC

In [9] the authors conducted an extensive study on the topology design problem to provide technical guidance to power utilities, for the design of the communication network for the PMU based real-time applications. Two principal architectural designs considered in [9] are shown in figs. 7.1 and 7.2 The difference between these two designs is that in the first case, the PMU data goes directly to the CC (fig. 7.1), and in the second case, it goes indirectly to the CC through a LCC (fig. 7.2). The results of our study of the first case are already presented in Chapter 6 [45], and results for the second case is presented in this chapter. Prior works have studied the optimal PMU placement problem with specific objectives, such as, full network observability [44] or joint optimization of PMU placement and associated communication infrastructure [46], [10]. The RDCMST problem was studied by most researchers in a topological setting [36], [37], whereas it is studied in a geometric setting in Chapter 6 [45].

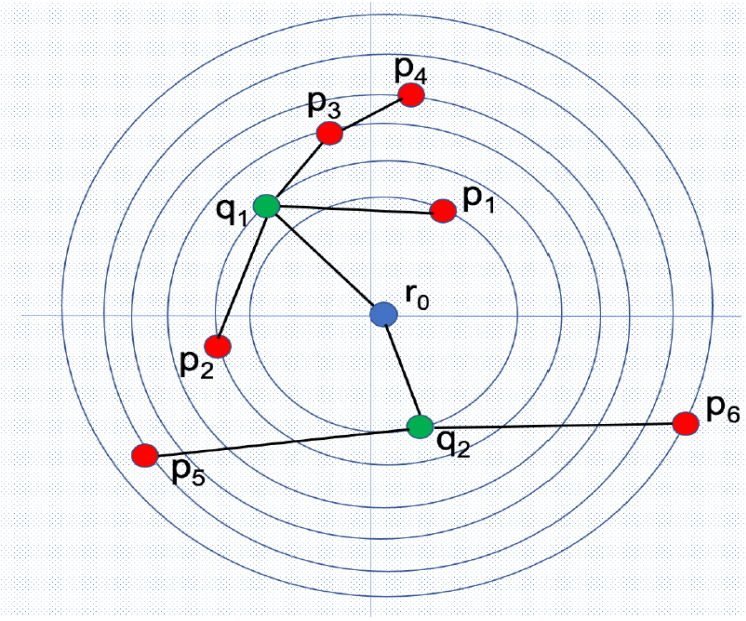

Fig. 7.3. A Possible Solution for the MRDCMSF Problem

In the abstract model of the multiple LCC scenario, a set of points $P = \{p_1, \dots, p_n\}$ representing the locations of the substations, a set of points $Q = \{q_1, \dots, q_m\}$ representing the locations of the LCCs, and a point r_0 representing the locations of the CC, can be considered. A layout of the substations with $n = 6$, the LCCs with $m = 2$ and the CC is shown in fig. 7.3. In this model, the PMU data from substations has to reach one LCC within an acceptable time threshold δ_1. More specifically, data from each point $p_i \in P$ must reach one point $q_j \in Q$ within the acceptable time threshold δ_1. The communication infrastructure essentially has to build a Multi-Rooted Delay Constrained Minimum Spanning Forest (MRDCMSF), where the roots correspond to the LCCs, and other nodes correspond to the SSs.

7.1. Problem Formulation

The inputs for the MRDCMSF problem are two sets of points $P = \{p_1, \dots, p_n\}$ and $Q = \{q_1, \dots, q_m\}$ representing the locations of the substations and the LCCs respectively, and a point r_0 representing the locations of the CC. From the input of the MRDCMSF problem, a weighted, complete graph $G = (V \cup U \cup \{R\}, E)$ is created, where each node $v_i \in V$ corresponds to a point $p_i \in P$ each node $u_j \in U$ corresponds to a point $q_j \in Q$ and the single node in the set R corresponds to the point r_0. The weights of the edges $w(v_i, v_j)$, $w(u_i, v_j)$ and $w(u_i, u_j)$ of an edge $e \in E$, is set equal to the Euclidean distance between the corresponding points. It is assumed that the weight $w(e)$ of an edge $e \in E$ represents both the cost and the delay of the link e. It is noted as a simplification, as in a realistic communication system, the link length is only one of the factors that will determine the

cost and the delay associated with that link. However, this simplified model is widely used in literature on communications networks as many efficient algorithms for more complicated models are based on efficient algorithms for this simplified model [38].

The objective of the MRDCMSF problem is to construct the least cost network, subject to the constraint that the length of the path from every node v_i corresponding to a point $p_i \in P$ to one node u_j corresponding to point $q_j \in Q$ does not exceed a specified delay threshold δ. This corresponds to the construction of the minimum cost spanning forest for the graph $G = (V, E)$, where the roots of the trees correspond to the points in the set Q and the other nodes correspond to the points in the set P, subject to the constraint that in the forest every node v_i ($p_i \in P$) has a path to a node u_j ($q_j \in Q$) of length at most δ.

7.2. Optimal Solution for MRDCMSF Problem

In this section, the Integer Linear Programming formulation of the MRDCMSF problem. First, a few notations are defined as follows:

$$f^{v_i \to u_j}: Flow\ from\ v_i\ to\ u_j, where\ v_i \in V_P\ and\ u_j \in V_Q$$

$$f_{(v_k,v_l)}^{v_i \to u_j} = \begin{cases} 1, & if\ flow\ from\ v_i\ to\ u_j\ uses\ the\ edge\ between\ nodes\ (v_k, v_l) \\ 0, & otherwise \end{cases} \quad (7.1)$$

F_{v_k,v_l}: $Total\ number\ of\ flows\ over\ the\ edge\ (v_k, v_l) from\ some\ v_i \in V_P\ to\ some\ u_j \in V_Q$.

$$F_{(v_k,v_l)} = \sum_{v_i \in V_P} \sum_{u_j \in V_Q} f_{(v_k,v_l)}^{v_i \to u_j} \quad (7.2)$$

It may be noted that while computing the total network design cost, design cost, the cost of edge (v_k, v_l) is counted only once, if total flow on the edge (v_k, v_l) exceeds zero, i.e., $F_{(v_k,v_l)}$>0. Accordingly, first a new binary variable $X_{(v_k,v_l)}$ is defined and then a constraint is introduced using this variable to ensure that the cost of the (v_k, v_l) edge (C_{v_k,v_l}) is counted only once when the total flow on that edge is greater than zero. The variable and the constraint are described next.

$$X_{(v_k,v_l)} = \begin{cases} 1, & if\ F_{(v_k,v_l)} \geq 1 \\ 0, & otherwise \end{cases} \tag{7.3}$$

$$F_{(v_k,v_l)} \leq (n-1)X_{(v_k,v_l)} \tag{7.4}$$

Next, the objective function of the MLCC problem is stated as:

$$Minimize \sum_{v_k \in V_P, v_l \in V_P \cup V_Q} C_{(v_k,v_l)} X_{(v_k,v_l)} \tag{7.5}$$

To establish a path from v_i to u_j, one must satisfy the flow constraints, i.e., the difference between incoming and outgoing flows at any node v_k must be equal to -1, +1, and 0, if v_k is the source, destination, or an intermediate node respectively. The constraints are stated as follows:

i. Constraint at source node v_i for the flow $f^{v_i \to u_j}$

$$\sum_{v_k \in N(v_i)} f_{(v_k,v_i)}^{v_i \to u_j} - \sum_{v_k \in N(v_i)} f_{(v_i,v_k)}^{v_i \to u_j} = -1 \tag{7.6}$$

ii. Constraint at destination node u_j for the flow $f^{v_i \to u_j}$

$$\sum_{v_k \in N(u_j)} f_{(v_k,u_j)}^{v_i \to u_j} - \sum_{v_k \in N(u_j)} f_{(u_j,v_k)}^{v_i \to u_j} = 1 \tag{7.7}$$

iii. $\forall\, v_l \in \{V_P \cup V_Q\} \backslash \{v_i, u_j\}$, constraint at intermediate nodes v_l for the flow $f^{v_i \to u_j}$

$$\sum_{v_k \in N(v_l)} f_{(v_k,v_l)}^{v_i \to u_j} - \sum_{v_k \in N(v_l)} f_{(v_l,v_k)}^{v_i \to u_j} = 0 \tag{7.8}$$

iv. Delay Constraint for the flow $f^{v_i \to u_j}$

$$\sum_{v_k, v_l \in V_P \cup V_Q} D_{(v_k,v_l)} f_{(v_k,v_l)}^{v_i \to u_j} \leq \delta \tag{7.9}$$

A solution that satisfies constraints 7.6, 7.7, 7.8 and 7.9 will establish a path from v_i to u_j that will satisfy the acceptable delay threshold δ. These four constraints may be combined to express in the canonical form of a constraint matrix $\mathcal{A}_{(v_i,u_j)}$ in the following way:

$$\mathcal{A}_{(v_i,u_j)} \mathcal{F}_{(v_k,v_l)}^{v_i \to u_j} \leq \mathcal{B}_{v_i,u_j} \tag{7.10}$$

It may be recalled that, starting was done with the set of *P* and *Q* representing the locations of the sub-stations and the local control centers. The constraints expressed in 10 relates only to the flow from $v_i \in V_P$ and $u_j \in V_Q$. It may also be recalled that every point $v_i \in V_P$ must send data to at least one $u_j \in V_Q$ within the delay threshold δ. Accordingly, as long as the delay on at least one path among the set of *m* paths ($m = |U|$), $v_i \to u_1$, $v_i \to u_2, \ldots, v_i \to u_m$ does not exceed the delay threshold δ the network design will satisfy its specification. This can be stated in terms of the canonical form in the following way:

$$\begin{aligned}
\mathcal{A}_{(v_i,u_1)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_1} &\le \mathcal{B}_{v_i,u_1} \\
&OR \\
\mathcal{A}_{(v_i,u_2)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_2} &\le \mathcal{B}_{v_i,u_2} \\
\dots \quad &OR \quad \dots \\
\dots \quad &OR \quad \dots \\
\dots \quad &OR \quad \dots \\
\mathcal{A}_{(v_i,u_m)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_m} &\le \mathcal{B}_{v_i,u_m}
\end{aligned} \tag{7.11}$$

The logical condition expressed in disjunctive form in (7.11) can be replaced by a set of linear equations with the introduction of a m binary variables $y_1, \dots, y_{m,}$ and a new constraint.

$$\sum_{i=1}^{m} y_i = 1 \tag{7.12}$$

$$\begin{aligned}
\mathcal{A}_{(v_i,u_1)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_1} &\le \mathcal{B}_{v_i,u_1} + M(1-y_1) \\
\mathcal{A}_{(v_i,u_2)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_2} &\le \mathcal{B}_{v_i,u_2} + M(1-y_2) \\
\dots \quad &\dots \quad \dots \\
\dots \quad &\dots \quad \dots \\
\dots \quad &\dots \quad \dots \\
\mathcal{A}_{(v_i,u_m)}\mathcal{F}_{(v_k,v_l)}^{v_i\to u_m} &\le \mathcal{B}_{v_i,u_m} + M(1-y_m)
\end{aligned} \tag{7.13}$$

where M is large constant.

Now the optimal solution to the MRDCMSF problem can be found by the solving the Integer Linear Program with the objective function expressed in eq. (7.5), subject to the constraints expressed in eqs. (7.2), (7.4), (7.12) and (7.13).

7.3. Lagrangian Relaxation Based Solution

In this section, to simplify the notations for a path formulation, the MRDCMSF problem is denoted by (G, c, d, δ), where $G = (V, E)$ is a graph representing an extended communication network with a candidate set of LCCs and communication links; The node set, V , consists of a CC, denoted by 0, a set of m candidate LCCs, denoted by $J := \{1, \dots, m\}$, and a set of n substations, denoted by $K \coloneqq \{m+1, \dots, m+n\}$. The adjacency of each pair of nodes in V is represented by a set of ordered pairs: $E \subseteq \{(i,j): i < j \ \in N\}$. In this section, it is assumed that an edge set with a special property that the CC is only connected to LCCs and LCCs are not adjacent to each other. For each edge $(i,j) \in E$, there exist a cost c_{ij} and a delay d_{ij} associated to it; δ denotes the delay threshold imposed on each path connecting each substation to the CC. To allow for a path formulation, G is converted into a directed graph by replacing each edge $(i,j) \in E$ for which $i \notin \{0\} \cup J$ with two arcs $(i,j), (j,i)$ and each edge $(i,j) \in E$ for which $i \in \{0\} \cup J$ with (i,j) and associate each arc with associated c_{ij} and d_{ij}. The resultant set is denoted by A.

7.3.1. A Path Formulation for the MLCC

In this section, the path formulation of the RDCMSTP in [47] is extended with an additional binary variable y_j associated to each LCC j, which takes on 1 if LCC j is active, and 0 otherwise.

$$min \sum_{(i,j)\in A:\ i\neq 0} c_{ij}X_{ij} + \sum_{j=1}^{m} c_{0j}y_j \qquad (7.14a)$$

$$s.t. \sum_{(i,k)\in A:\, i\neq 0} X_{ik} = 1\,, \forall k \in K, \tag{7.14b}$$

$$X_{jk} \leq y_j, (j,k) \in A, \tag{7.14c}$$

$$\sum_{p\in P^k} U_p^k = 1, \forall k \in K, \tag{7.14d}$$

$$\sum_{p\in P_{ij}^k} U_p^k \leq X_{ij}, \qquad \forall (i,j) \in A, i \neq 0, k \in K, \tag{7.14e}$$

$$\sum_{p\in P_{0j}^k} U_p^k \leq y_j\,, \quad \forall j \in J, k \in K, \tag{7.14f}$$

$$y_j \in \{0,1\}, \qquad \forall j \in J, \tag{7.14g}$$

$$X_{ij} \in \{0,1\}, \qquad \forall (i,j) \in A, i \neq 0, \tag{7.14h}$$

$$U_p^k \in \{0,1\}, \qquad \forall p \in P^k, k \in K \tag{7.14i}$$

It is assumed that the set of all paths connecting the CC to each substation *k* within the delay threshold is available and it is denoted here by P^k. In addition, $P_{ij}^k \subseteq P^k$ denotes the set of paths having the arc $(i,j) \in A$ among those in P^k. Since each substation must be connected to the CC, there should be at least one path available in P^k for each substation *k*. This requirement is formulated by introducing a binary variable U_p^k for each $k \in K$ and $p \in P^k$, which takes on 1 if the path *p* is used to connect the CC to the substation *k*, and 0 otherwise. Using these path binary variables, the requirement can be expressed as in eq. (7.14d). Moreover, for a path *p* to be available, all arcs comprising the path should be

available. The availability of each arc $(i,j) \in A$ with $i \neq 0$, is modelled by introducing a binary variable X_{ij}, which takes on 1 if the arc (i,j) is available, and 0 otherwise. Then, eq. (7.14e) and (7.14f) will enforce that a path with an arc (i,j) can be active only when the arc (i,j) is available $i.e., \big(X_{ij} = 1\ for\ (i,j) \in A\ with\ i \neq 0\ or\ y_j = 1\ for\ (i,j) \in A\ with\ i = 0\big)$. Eq. (7.14c) constrains that a downstream arc connected to a LCC j can be active only when the LCC is installed. Eq. (7.14b) requires that each substation has exactly one active incoming arc.

7.3.2. Lagrangian Dual

If we relax Constraints (7.14e)-(7.14f), the problem is decomposed into a variant of a uncapacitated facility location problem and a constrained shortest path problem, each of which has a rich literature that we can leverage. To be specific, let λ_{ij}^k be the dual variable associated with Constraint (7.14e) and μ_{ij}^k be that of Constraint (7.14f). Then, the resultant Lagrangian dual is as follows:

$$max_{\lambda \geq 0, \mu \geq 0} \big(UFL_{(\lambda,\mu)}\big) + \sum\nolimits_{k \in K} \big(CSP_{(\lambda,\mu)}^k\big),$$

where $\big(UFL_{(\lambda,\mu)}\big) \coloneqq$

$$min \sum_{(i,j) \in A : i \neq 0} \left(c_{ij} - \sum_{k \in K} \lambda_{ij}^k\right) X_{ij} + \sum_{j \in J} (c_{0j} - \sum_{k \in K} \mu_j^k) y_j \qquad (7.15a)$$

$$s.t.\ (7.14b), (7.14c), (7.14g), and\ (7.14h), \qquad (7.15b)$$

and $\big(CSP_{(\lambda,\mu)}^k\big) \coloneqq$

$$min \sum_{(i,j)\in A: i\neq 0} \lambda_{ij}^k \sum_{p\in P_{ij}^k} U_p^k + \sum_{j=1}^{m} \mu_j^k \sum_{p\in P_{0j}^k} U_p^k, \tag{7.16a}$$

$$s.t. (7.14d)\ and\ (7.14i), \tag{7.16b}$$

which is equivalent to the following constrained shortest path problem:

$$min \sum_{(i,j)\in A: i\neq 0} \lambda_{ij}^k x_{ij}^k + \sum_{j\in J} \mu_j^k x_{0j}^k \tag{7.17a}$$

$$s.t. \sum_{(0,j)\in A} x_{0j}^k = 1, \tag{7.17b}$$

$$\sum_{(i,j)} x_{ij}^k - \sum_{(j,i)\in A} x_{ji}^k = 0, \qquad \forall j \in J \cup K: j \neq k, \tag{7.17c}$$

$$\sum_{(k,j)\in A} x_{kj}^k - \sum_{(j,k)\in A} x_{jk}^k = -1, \tag{7.17d}$$

$$\sum_{(i,j)\in A} d_{ij} x_{ij}^k \leq \delta, \tag{7.17e}$$

$$x_{ij}^k \in \{0,1\}, \qquad \forall (i,j) \in A. \tag{7.17f}$$

7.3.3. Subgradient Method for Solving the Lagrangian Dual

It is to be noted that the Lagrangian dual problem is a concave maximization problem in λ and μ, since $L(\lambda,\mu) \coloneqq \left(UFL_{(\lambda,\mu)}\right) + \sum_{k=m}^{m+n}(CSP_{(\lambda,\mu)}^k)$ is a pointwise minimum of affine functions in λ and μ. Therefore, a subgradient method can be used for solving the Lagrangian dual. The subgradient method starts from an initial point (λ^0, μ^0)

and iteratively moves toward the global maximum of L while being guided by its subgradient at the current point.

Algorithm 7.1: Subgradient method

$l \leftarrow 0$;

$\lambda, \mu \leftarrow \lambda^0, \mu^0$;

$UB \leftarrow \infty$;

while true **do**

Compute $L(\lambda^l, \mu^l)$ and obtain its solution $\widehat{X_{ij}}$, $\widehat{y_j}$, and $\widehat{x_{ij}^k}$;

Compute a subgradient g^l of the function L at (λ^l, μ^l) as in Eq. (7.18);

if $\|g^l\| < \epsilon$ **then**

Stop, the approximated optimal objective value is $L(\lambda^l, \mu^l)$;

Compute UB_{temp} using $\widehat{x_{ij}^k}$;

$UB \leftarrow min\{UB, UB_{temp}\}$;

Compute $(\lambda^{l+1}, \mu^{l+1}) = [(\lambda^l, \mu^l) + \alpha^l g^l]_+$, where α^l is the stepsize at this step, computed as in Eq. (7.19), and $[\,.\,]_+$ means a projection onto the nonnegative orthant $(e.g., \ \ ([a]_+)_i = a_i \ \ if \ a_i \geq 0, and \ 0 \ otherwise)$;

$l \leftarrow l + 1$;

A subgradient of L at (λ, μ) can be obtained as follows:

$$\frac{\partial L(\lambda, \mu)}{\partial \lambda_{ij}^k} = -\sum_{(i,j) \in A: i \neq 0} \widehat{X_{ij}} + \widehat{x_{ij}^k}, \qquad \forall k \in K, (i,j) \in A, i \neq 0, \qquad (7.18a)$$

$$\frac{\partial L(\lambda, \mu)}{\partial \mu_{ij}^k} = -\sum_{j \in J}^{m} \widehat{y_j} + \widehat{x_{0j}^k}, \qquad \forall k \in K, j \in J, \qquad (7.18b)$$

where $\widehat{X_{ij}}$, $\widehat{y_j}$, and $\widehat{x_{ij}^k}$ respectively correspond to the optimal solution of $\left(UFL_{(\lambda,\mu)}\right)$ and $\left(CSP_{(\lambda,\mu)}^k\right)$. Using the shortest paths obtained when solving $\left(CSP_{(\lambda,\mu)}^k\right)$ for each

iteration, we can compute the upper bound, and the computed $L(\lambda, \mu)$ will serve as a lower bound. In practice, the stepsize is updated using the best available upper bound, denoted by UB, as follows:

$$\alpha = \frac{UB - L(\lambda, \mu)}{\|v\|^2}, \tag{7.19}$$

where v is the vector describing the violation of each relaxed constraint for (λ, μ). The overall procedure is described in Algorithm 7.1.

7.4. Heuristic Solution for the MRDCMSF Problem

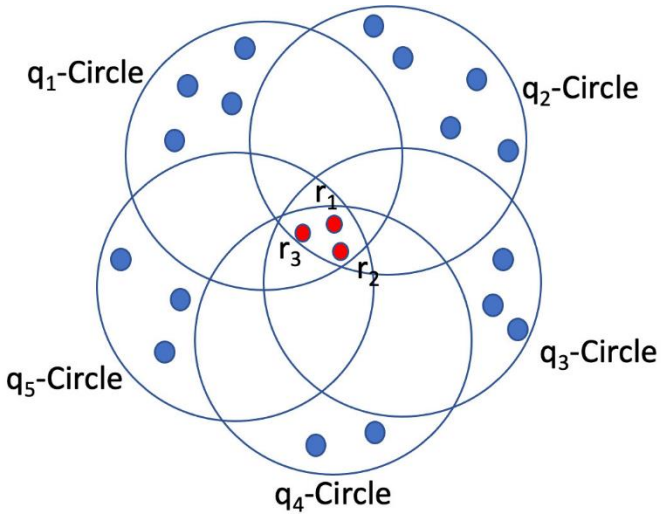

Fig. 7.4. Substations (red points) in the intersection (contention) area of several q-circles, where each q-circle correspond to an LCC

A MRDCMSF problem instance has a number of LCCs and as long as PMU data from a substation arrives at one of the LCCs within the specified delay threshold δ, the delay constraint is satisfied. Accordingly, the MRDCMSF problem is significantly more complex, because each substation has to decide which LCC to send the PMU data to

without violating the delay constraint and minimizing the network design cost. In other words, the set of substations has to be partitioned into groups, where each group will be associated with one LCC. This partitioning of the substations into groups with a goal to minimize the design cost, without violating the delay constraint, makes the MRDCMSF problem challenging.

Algorithm 7.2: Contention Resolution Algorithm

Input: A set of $P = \{p_1, \dots, p_n\}$, $Q = \{q_1, \dots, q_m\}$ points and delay threshold δ.

Output: Partitioning of the set *P* into *m* disjoint subsets $\{P_1, \dots, P_m\}$

1. Corresponding to every point $p_i, 1 \leq i \leq n$ and $q_i, 1 \leq i \leq m$ create two sets called $PSet(p_i)$ and $QSet(q_i)$ and initialize both sets to be empty

2. **for** $i = 1$ **to** n **do**
   for $j = 1$ **to** m **do**
   Compute distance $d(p_i, q_j)$ between points p_i and q_j
   if $d(p_i, q_j) \leq \delta$; then
   $PSet(p_i) = PSet(p_i) \cup \{q_j\}$

3. Identify those $PSet(p_i)s$, whose $|PSet(p_i)| = 1$, and denote this set as $S = \{S_1, \dots, S_k\}$. It may be noted that each $S_i, 1 \leq i \leq k$ contains exactly one element of *Q*.

4. **for** $i = 1$ to *m* **do**
   for $j = 1$ to *k* **do**
   if $q_k \in S_j = PSet(p_l)$ then
   $Qset(Q_k) = Qset(Q_k) \cup p_l$

5. $P' = \cup_{r=1}^{k} QSet(Q_r)$ (P' is a subset of *P* points that are contention-free, in the sense that only one *q* point can be reached from these *p* points within the specified delay threshold δ.)

6. Construct the RDCMST for these contention free *p* points with the corresponding *q* point as the root. Let these trees be $T_1, \dots, T_s$.

7. $P'' = P - P'$ (P'' is the subset of P points that are involved in contention, in the sense that multiple q point can be reached from all these p points within the specified delay threshold δ).

8. Suppose that $p_i \in P''$ and points q_x, q_y, q_z can be reached from p_i within the delay threshold δ. In step 6, RDCMST is constructed with roots q_x, q_y, q_z and those p points that were contention-free. Suppose that the cost of these RDCMSTs with roots q_x, q_y, q_z are $C(T_x), C(T_y), C(T_z)$ respectively.

9. Compute the cost of the RDCMST of the trees with roots q_x, q_y, q_z and point (node) set $T_x \cup \{p_i\}, T_y \cup \{p_i\}, T_z \cup \{p_i\}$. The cost of these trees are $C(T_x \cup \{p_i\})$, $C(T_y \cup \{p_i\})$, $C(T_z \cup \{p_i\})$ respectively.

10. Compute the marginal cost of adding the point p_i to the RDCMSTs with roots q_x, q_y, q_z as:

    $C(T_x \cup \{p_i\}) - C(T_x)$,
    $C(T_y \cup \{p_i\}) - C(T_y)$ and
    $C(T_z \cup \{p_i\}) - C(T_z)$ respectively

11. Assign p_i to be a part of that RDCMST (with roots q_x, q_y, q_z) whose marginal cost of adding p_i is the smallest.

In the notation of this chapter, P and Q are the sets of points that represent the locations of the substations and LCCs respectively. In this model both cost and delay on a link connecting any two points is equal to the Euclidean distance between the points. If the Euclidean distance between two points $p_i \in P$ and a $q_j \in Q$ is greater than the delay threshold δ, there is no way the substation at p_i can deliver PMU data to q_j within the delay threshold δ. In fig. 7.4. the blue and red points represent the locations of the substations. The blue points represent the locations of the substations which are within the q circle of only one LCC and hence there is no question as to which LCC they should send their PMU data to. The red points represent the locations of the substations which are within the q circles of multiple LCCs and hence there is question as to which LCC they should

send their PMU data. One has to resolve the contention of the red points in fig. 7.4. to decide which q circle a specific red point should be grouped with. Once the contention is resolved, the problem becomes same as the RDCMST problem which can be solved by the M-Prim algorithm [45] . The contention resolution algorithm (Algorithm 7.2.) presented in this section uses the M_Prim algorithm (Algorithm 6.1.) presented in Chapter 6.

Algorithm 7.2. resolves contentions by first constructing a collection of spanning trees with the LCCs as the roots with the contention-free p points. Suppose $T_1, \dots, T_s$ are the trees constructed in this process, and that i-substation (point p_i) is in the intersection area of q-circles of points q_x, q_y, q_z. The decision to assign p_i to $q_x, q_y \; or \; q_z$ is made by first computing the marginal cost of adding the point p_i to the trees $T_x, T_y \; and \; T_z$ corresponding to $q_x, q_y \; and \; q_z$ constructed earlier, and then assigning it to the tree whose the marginal cost increase is minimum. Algorithm 7.2 describes the process in detail. Once the contentions are resolved, a collection of spanning trees are constructed using the M-Prim algorithm given in Chapter 6.

7.5. Experimental Results

The performance of all the solution techniques presented in this chapter is evaluated with substations located in Tucson and Phoenix. The latitude-longitude locations of substations in Arizona are obtained from the U.S. Dept. of Homeland Security website [43] and are shown in fig. 7.3.

The blue and red dots indicate the locations of operational/non-operational substations respectively. The total number of substations in Arizona is 892, of which 653

are operational. The number of substations in Phoenix and Tucson are 132 and 25 respectively. As PMU installation in only 20%-30% of the substations is sufficient for full observability [44], this information is used in the experiments done in this chapter.

Data Sets	#LCC	$\delta = 50$		$\delta = 60$		$\delta = 70$		$\delta = 80$	
		H/OP	LUB/OP	H/OP	LUB/OP	H/OP	LUB/OP	H/OP	LUB/OP
DS-1 (12 PMUs)	2	1.008	1.05	1.007	1.02	1.007	1.003	1	1
	3	1	1.004	1	1.002	1	1.002	1	1
	4	1	1.003	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1
DS-2 (11 PMUs)	2	1.002	1.013	1	1.005	1	1.002	1	1
	3	1	1.002	1	1.005	1	1	1	1
	4	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1
DS-3 (10 PMUs)	2	1.01	1.04	1	1.008	1	1.007	1	1
	3	1	1.008	1	1.004	1	1	1	1
	4	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1
DS-4 (8 PMUs)	2	1	1.002	1	1.001	1	1	1	1
	3	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1	1
DS-5 (4 PMUs)	2	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1

Table 7.1. H/Op And LUB/Op: Ratios Between The Heuristic And The Optimal; And Lagrangian Upper Bound And The Optimal, Respectively For The Tucson Data Set

A part of the results of the experiments are presented in Table 7.1. and Table 7.2. The delay threshold δ was varied from 50 to 80. Thirty instances of each data set (DS-1 through DS-5) were created and results with ILP, Lagrangian and Heuristic were computed. The average of the results of thirty instances for a specific δ value is reported in the Table 7.1. The ratios between the costs of the Heuristic and the Optimal and the Upper Bound of Lagrangian to Optimal, with variation of δ values from 50 to 80 is reported in Table 7.1 and 7.2.

Data Sets	No. of LCC	$\delta = 50$		$\delta = 60$		$\delta = 70$		$\delta = 80$	
		H/OP	LUB/OP	H/OP	LUB/OP	H/OP	LUB/OP	H/OP	LUB/OP
DS-1 (53 PMUs)	5	1.06	1.17	1.03	1.03	1	1.03	1	1.006
	8	1.04	1.13	1.02	1.04	1	1.02	1	1.0001
	10	1	1.06	1	1.03	1	1.01	1	1.0002
	12	1	1.02	1	1.02	1	1	1	1
	15	1	1.01	1	1.01	1	1	1	1
DS-2 (44 PMUs)	5	1.06	1.07	1.05	1.03	1	1.02	1	1.005
	8	1.02	1.07	1.01	1.02	1	1.02	1	1.002
	10	1	1.02	1	1.02	1	1.0001	1	1.0001
	12	1	1.0003	1	1	1	1	1	1
	15	1	1	1	1	1	1	1	1
DS-3 (40 PMUs)	5	1.03	1.07	1.04	1.06	1	1.02	1	1.004
	8	1	1.06	1	1.04	1	1.02	1	1.0002
	10	1	1.07	1	1.03	1	1.0005	1	1
	12	1	1.07	1	1.03	1	1	1	1
	15	1	1.04	1	1	1	1	1	1
DS-4 (33 PMUs)	5	1.006	1.09	1	1.06	1	1.05	1	1.006
	8	1	1.04	1	1.02	1	1.002	1	1
	10	1	1	1	1.0003	1	1	1	1
	12	1	1	1	1	1	1	1	1
DS-5 (26 PMUs)	5	1.02	1.003	1	1.002	1	1	1	1
	8	1	1.0007	1	1.0003	1	1	1	1
	10	1	1	1	1	1	1	1	1
	12	1	1	1	1	1	1	1	1

Table 7.2. H/Op And LUB/Op: Ratios Between The Heuristic And The Optimal; And Lagrangian Upper Bound And The Optimal, Respectively For The Phoenix Data Set

It can be seen in Table 7.1 and 7.2 that both the Heuristic and the Lagrangian solution produce either optimal or near optimal solution for all problem instances. Significantly more experiments are conducted with much larger number of PMUs and LCCs, a part of them for Tucson city of Arizona is presented in Table 7.1. The other table Table 7.2. shows the results for Phoenix city. As the maximum deviation from the optimal for the Heuristic and the Lagrangian were 6% and 7% respectively in case of Tucson data, it can be inferred that both of techniques produce high quality solution. Computational cost of the techniques was also comparable and quite reasonable. The problem instances are divided into three classes−small, medium and large depending on the number of PMUs in

the problem instances. For data sets of both cities, the instances with 4 to 20, 21 to 35 and 36 to 53 PMUs were considered small, medium and large respectively. The computation time for the ILP varied from 1-5, 38-224 and 412-818 seconds for the small, medium and large data sets respectively. For the Lagrangian relaxation, the corresponding numbers were 3-60, 62-492 and 635-1224 seconds. The heuristics produced solution for all three classes in less than one second. From the experiments it can be concluded that almost always the heuristic produces a high-quality solution at a low computational cost.

Chapter 8

JOINT PMU PLACEMENT AND OPTIMAL COST NETWORK DESIGN FOR BOUNDED DELAY DATA TRANSFER FROM SUBSTATIONS TO CONTROL CENTERS

There has been extensive number of studies on optimal PMU placement problem [12]. Communication network topology design for the smart grid environment also has received attention from the researchers. Most of the studies in optimal PMU placement and communication network topology was done in isolation. Only in the recent years, researchers have pointed out that conducting PMU placement and communication network topology design in isolation, may not really lead to total infrastructure design cost minimization [46]. In order to minimize total infrastructure design cost, both PMU placement and topology design problems must be considered simultaneously.

8.1. Joint PMU Placement and Communication Network Topology Design (JPMUPCNTD) Problem

In this section, the joint design of PMU placement and communication topology is studied. The need for joint design instead of design in isolation is elaborated here with the help of an example. In fig. 8.1(a), three substations located at points p_1, p_2 and p_3 and a substation. located at q_1 is shown. The bus in p_1 is connected to the busses in p_2 and p_3 through transmission lines p_2 and p_3 through transmission lines (shown as black solid lines). For complete observability it's sufficient to place only one PMU at p_1 (shown as a

square box in fig. 8.1(b). This this the optimal solution for the PMU placement problem if it's considered in isolation. The data from the PMU at p_1 must arrive at the control center at q_1 within the specified delay threshold. The least expensive way to achieve that may be a direct connection (say, a fiber optic line) between the points p_1 and q_1. If we denote the cost of a PMU as C_{PMU} and the cost of fiber optic line of length d_1(distance between the points p_1 and q_1) as $C_{fiber}(d_1)$, the total cost of this design will be $C_{PMU} + C_{fiber}(d_1)$. The substations would have been completely observable, if instead of placing a PMU at p_1, two PMUs are placed at p_2 and p_3 respectively, as shown in fig. 8.1(c), the cost of this design is $2 \times C_{PMU} + C_{fiber}(d_1) + C_{fiber}(d_2)$. Although the second design may not be optimal from the PMU placement problem, if considered in isolation (it uses two PMUs instead of one), the design may be optimal from joint design perspective, if $C_{PMU} < C_{fiber}(d1) - C_{fiber}(d2) - C_{fiber}(d3)$. For this reason, in this section a joint design is focused on.

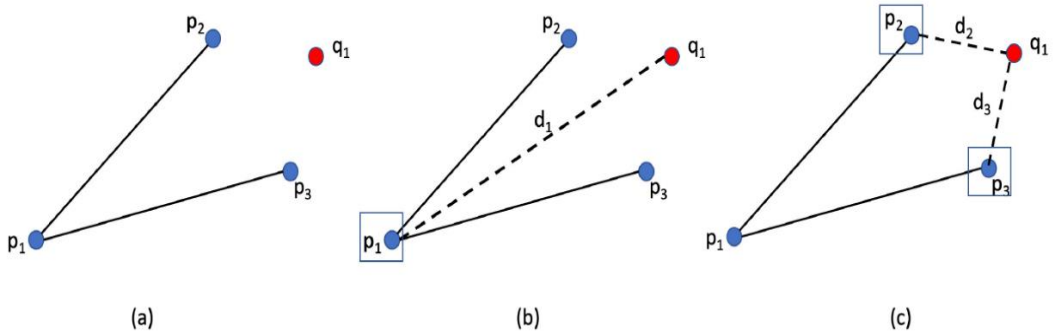

Fig. 8.1. Network Connecting Substations and Central Control Center(s)

8.2. Optimal Solution for JPMUPCNTD Problem

In Chapter 6 of this dissertation, the ILP for the RDCMST problem is presented when the locations of the PMUs have already been determined and are known. The difference between the RDCMST and JPMUPCNTD is that while in RDCMST the

locations of the PMUs are known, in JPMUPCNTD they are not known and have to be determined as a part of finding the optimal solution.

It may be recalled that in Chapter 6, the delay from every point where a PMU is placed to the root node had to be at most δ, expressed in the following way:

i. Delay Constraint: $\forall k, 2 \leq k \leq n, \sum_{i=1}^{n} \sum_{j=1}^{n} D_{i,j} f_{i,j} \leq \delta$ (8.1)

It may be further recalled the objective function in Chapter 6 was stated as:

ii. Objective Function: $Minimize \ \sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j}^{link} X_{i,j}$ (8.2)

and the other constraints were stated as follows:

iii. Constraint at node v_1 for the flow f^k from v_k to v_1:

$$\forall \, k, 2 \leq k \leq n, \sum_{j \in N(v_1)} f_{j,1}^{k} - \sum_{j \in N(v_1)} f_{1,j}^{k} = 1 \tag{8.3}$$

iv. Constraint at node v_k for the flow f^k from v_k to v_1:

$$\forall \, k, 2 \leq k \leq n, \sum_{j \in N(v_1)} f_{j,k}^{k} - \sum_{j \in N(v_1)} f_{k,j}^{k} = -1 \tag{8.4}$$

v. Constraints at node v_l, $\left(v_l \in \{V - \{1, k\}\}\right)$ for the flow f^k from $v_k \longrightarrow v_1$:

$$\forall \, k, 2 \leq k \leq n, \sum_{j \in N(v_l)} f_{j,l}^{k} - \sum_{j \in N(v_l)} f_{l,j}^{k} = 0 \tag{8.5}$$

When joint PMU placement and communication network design is considered, the originating point of the *k*-th flow is unknown and has to be found out. For complete observability of the electric power grid, the PMUs must be placed at the subset of the nodes

in the topological graph of the electric power grid system that constitute the Dominating Set [40] of the corresponding graph. The IEEE 14-Bus system is considered here as an example and also for explaining how the topological graph is formed (fig.8.2). The substation division is done here in the same way as in Chapter 3. Therefore, the topological graph for the power layer of a smart grid formed using IEEE 14-Bus, can be given as in fig. 8.2.

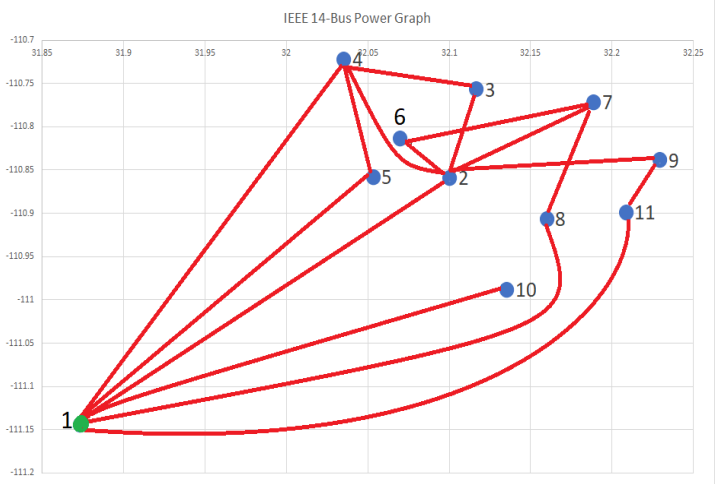

Fig. 8.2. Power Layer Graph for IEEE 14-Bus Smart Grid

In fig. 8.2. the substation locations are real substation locations of Arizona. Substation location data can be obtained from the U.S. Homeland security department website [43] but there is no bus information in that data. Therefore, in this chapter, the standard IEEE bus systems are superimposed on real substation locations of Arizona to test the ILP based solution to the problem. In order to do that, first a standard IEEE bus system, for example IEEE 14-Bus is selected and it is divided into a number of substations, following the same method as in Chapter 3. Now, these substations are given co-ordinates which match real substation location co-ordinates of Arizona. Since IEEE 14-Bus system has 11 SSs, in this chapter, random 11 substation locations are selected from Arizona and

each such substation is given a substation number $\{S_1, \dots, S_{11}\}$. Since both S_1 and S_2 have the same number of connections with other substations, any one can be selected as the CC. In this case, S_1 is selected. In fig. 8.2., S_1 is colored green to demarcate it as the CC. The power graph contains edges which represent the transmission lines connecting different substations in the standard IEEE bus systems. The substation S_1 is marked as node v_1 is this case and all the other substations are numbered from v_2 through v_{11}.

Now, the adjacency matrix of the topological graph of the electric power grid is taken as input to the JPMUPCNTD problem and it is denoted by A which is an $n \times n$ matrix, where n is the number of buses in the system. Let, $R = \{b_1, \dots, b_n\}$ is the set of all buses in the system. Each element of the matrix A can be defined as follows:

$$a_{i,j} = \begin{cases} 1 & if\ bus\ i\ and\ bus\ j\ are\ connected\ by\ a\ transmission\ line/transformer \\ 1 & if\ bus\ i = bus\ j \\ 0 & Otherwise \end{cases} \quad (8.6)$$

Other inputs to this problem include: an $n \times n$ cost matrix that stores the distances between every pair (i, j) of substations given as $C_{i,j}^{link}$ and cost of each PMU is taken a standard constant value given as C^{PMU}. The output of this problem is the total cost considering optimal PMU placement and the optimal cost tree connecting only the PMU containing SSs and the CC.

Now in order to formulate this problem using ILP, a binary variable z_k is taken, such that:

$$z_k = \begin{cases} 1, & if a\ PMU\ is\ placed\ at\ the\ bus\ in\ location\ b_k \\ 0, & otherwise \end{cases} \quad (8.7)$$

Suppose that Z is the vector comprises of all $z_k, 1 \leq k \leq n$. Then for full observability, one has to place PMUs in locations such that the following constraint is satisfied.

$$AZ^T \geq [1 \dots 1]^T \tag{8.8}$$

When we try to jointly optimize PMU placement and communication network design, we need to minimize the sum of the costs of PMUs and the communication links. Thus, the objective function should now be:

$$Minimize \sum_{i=1}^{n} C_i^{PMU} z_i + \sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j}^{link} X_{i,j} \tag{8.9}$$

Now, as mentioned in eq. 8.7, if the binary variable z_k gets a value 1, that indicates a PMU is placed at bus *k* or in other words at the substation containing bus *k*. This substation is denoted as v_k. So, the flow constraints for $z_k = 1$ can be stated as follows:

vi. If $z_k = 1$ then constraint at node v_1 for the flow f^k from v_k to v_1 is:

$$\sum_{j \in N(v_1)} f_{j,1}^k - \sum_{j \in N(v_1)} f_{1,j}^k = 1 \tag{8.10}$$

vii. If $z_k = 1$ then constraint at node v_k for the flow f^k from v_k to v_1 is:

$$\sum_{j \in N(v_k)} f_{j,k}^k - \sum_{j \in N(v_k)} f_{k,j}^k = -1 \tag{8.11}$$

viii. If $z_k = 1$ then the constraints for all nodes $v_t \in \{V - \{1, k\}\}$ for the flow f^k from v_k to v_1 is:

$$\sum_{j \in N(v_t)} f_{j,t}^k - \sum_{j \in N(v_t)} f_{t,j}^k = 0 \tag{8.12}$$

Now the flow constraints for joint PMU placement and communication network design problem can be stated as follows:

ix. If $z_k = 1$ then constraint at node u is:

$$Flow\_Diff_{(u)}^{(k)} = 1, \;\; if\; u = v_1 \tag{8.13}$$

x. If $z_k = 1$ then constraint at node u is:

$$Flow\_Diff_{(u)}^{(k)} = -1, \;\; if\; u = v_k \tag{8.14}$$

xi. If $z_k = 1$ then constraint at node u is:

$$Flow\_Diff_{(u)}^{(k)} = 0, \;\; if\; u \in \{V \setminus \{1, k\}\} \tag{8.15}$$

The logical statements in ix, x and xi can be expressed in terms of arithmetic equations needed for the ILP in the following way. Constraint (ix) can be replaced by two arithmetic inequalities as follows:

ix. a. $$Flow\_Diff_{(u)}^{(k)} \leq 1 + m(1 - z_k) \tag{8.16}$$

ix. b. $$Flow\_Diff_{(u)}^{(k)} \geq z_k \tag{8.17}$$

where m is the total number of edges in the topological graph of the electric power grid.

Constraint x can be replaced by two arithmetic inequalities as follows:

x. a. $$Flow\_Diff_{(u)}^{(k)} \leq -1 + m(1 - z_k) \tag{8.18}$$

x. b. $$Flow\_Diff_{(u)}^{(k)} \geq z_k - 2 \tag{8.19}$$

Similarly, constraint xi. can be replaced by two arithmetic inequalities as follows:

xi. a. $$Flow\_Diff_{(u)}^{(k)} \leq m(1 - z_k) \tag{8.20}$$

xi. b. $$Flow\_Diff_{(u)}^{(k)} \geq z_k - 1 \tag{8.21}$$

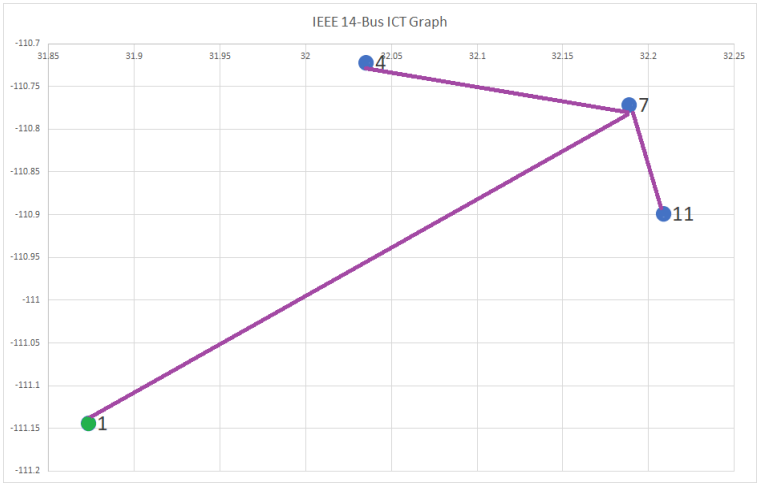

Fig. 8.3. ICT layer graph for IEEE 14-Bus smart grid

After the ILP runs on the smart grid system of IEEE 14-Bus described above, PMUs are placed at the following four bus locations for complete observability of the power grid: $\{P_2, P_7, P_{10}, P_{13}\}$. Here P_i denotes a bus with ID i. The corresponding substations with PMUs are: S_1 (containing bus P_7), S_4 (containing bus P_2), S_7 (containing bus P_{13}) and S_{11} (containing bus P_{10}). All the PMU containing substations should be connected to the CC in the final ICT design and in this case the CC S_1 is also a PMU containing substation. So, the final ICT design is given in fig. 8.3.

8.3. Experimental Results

It is already mentioned in section 8.2., the smart grids considered in this chapter are formed by superimposing standard IEEE Bus systems on real substation locations of Arizona. In this chapter the largest dataset considered is IEEE 300-Bus which can be

divided into 237 substations. The results for all the different datasets considered here are given in Table 8.1.

# Buses	# SSs	# PMUs	PMU Locations	#SSs with PMUs	Total PMU Cost	Total Link Cost	Total ICT Cost	Time in (s)
14	11	4	P_2, P_7, P_{10}, P_{13}	4	2000	458832	460832	0.012
30	26	10	$P_1, P_5, P_6, P_{10,} P_{12}, P_{15}, P_{19}, P_{25,} P_{27}$	8	5000	782654.39	787654.39	14.76
118	107	32	$P_1, P_5, P_9, P_{12}, P_{13}, P_{17}, P_{21}, P_{23} P_{26}, P_{28}, P_{34}, P_{37}, P_{41}, P_{45,} P_{49}, P_{53}, P_{56}, P_{62}, P_{63}, P_{68}, P_{71}, P_{75}, P_{77}, P_{80}, P_{85}, P_{86,} P_{90}, P_{94}, P_{101}, P_{105}, P_{110}, P_{114}$	32	16000	874273.66	890273.66	128.37
300	237	87	Cannot fit in the table	78	43500	1170121.92	1213621.92	421.39

Table 8.1. Joint PMU Placement and ICT Network Design Cost for Different Smart Grid Systems

The cost of a single PMU is considered here as $500, and the cost of communication links are considered as $1 per linear foot [48]. Cost of other ICT entities required for designing the whole network is not considered in this study.

Chapter 9

CONCLUSION AND FUTURE DIRECTIONS

It is to be concluded that, this dissertation aims to conduct an in-depth study in order to understand the interdependent critical infrastructures with focus on smart grids and tried to analyze and observe the vulnerabilities pertaining to such systems and proposed strategies to mitigate them. It should be noted that, an accurate estimation of the system's operating status in case of a failure, even before the failure occurs or even while that failure is manifesting, is extremely important and valuable. However, doing this for the modern smart grid is difficult because the intra-and-inter-dependencies between its power-and-communication networks are not well understood in the models surveyed in Chapter 2 of this dissertation. The model presented in Chapter 3, named Modified Implicative Interdependency Model (MIIM), is an effort by us to correctly capture these dependencies. The power system application that is used to demonstrate the practical utility of MIIM is state estimation. The results indicate that MIIM, in comparison to its predecessor, IIM, is more realistic in estimating the system state after some failure has occurred in the joint power-communication network. The future scope of work includes the use of MIIM for examining more complicated power/communication failure scenarios, such as EMP attacks, as well as for analyzing progressive recovery options of the joint network following a blackout/brownout.

Using a dual-platform based simulation of the smart grid system to verify the interdependency model MIIM is a novel approach proposed in Chapter 4 of this

dissertation. Since the MIIM ILP based solution is verified here using this co-simulation, the K-most vulnerable entities for any huge smart grid system can be easily obtained in a short time by using the ILP based solution considering the MIIM IDRs. Therefore, this approach can be used in real smart grids to obtain a self-updating K-Contingency list just by updating IDRs after any failure in the system. A scope of future work can be finding a method to use the real PMU data that is sent to the CCs at the rate of 30 samples per second to update the MIIM IDRs and automatically update the K-most vulnerable entities each time a new failure takes place in the system. Computing the ILP based contingency list within a very short period of time can be challenging and thus a suitable heuristic solution that takes into account all the observations of the simulation can also be a scope of future work. Also, the MIIM can be used to generate recovery list for the smart grid after a massive failure.

Securing the ICT network against cyber threats is another important aspect of this dissertation. The region based remote monitoring adopted by the Secure Smart Grid Monitoring Technique (SSGMT) presented in Chapter 5, helps in easy identification of an attack in the communication network of the smart grid. Security of sensor (RTU/PMU) data is of utmost importance in the smart grid as any alteration of the data can lead to wrong decision by the operator and this might badly harm the smart grid. SSGMT obtains data privacy by the encryption/decryption mechanisms, data integrity and authenticity by the one-way hash chain and HMACs respectively. Designing a threat model with attacks on smart sensors, gateways or servers and analyzing the effect of cyber-attacks on power grid can be evaluated in the future to offer more reliability on the ICT systems of a smart grid.A

further modification of the MIIM model to present the MSIIM model in Chapter 5 helps in even more accurate estimation of the operational levels of the ICT entities in the smart grid. This model will also help the operators to identify a false data injection and prevent that data from getting delivered to the destination PDC. Backtracking to the source of the attack is another feature offered by this model which helps in easy isolation of a compromised ICT entity in the system. In the proposed work PMUs are considered to be trusted. Identifying false data injection by PMUs using the MSIIM model and thereby taking necessary actions can be accomplished as a future work.

Other than modeling the interdependencies and providing security the ICT network for smart grid, this dissertation has also focused on building an optimal cost ICT network for the grid. Even though, the M_Prim algorithm for communication topology design presented in Chapter 6, does not guarantee to produce an optimal solution for all problem instances and in fact it's performance can be bad in some pathological case, the experimental evaluations performed in Chapter 6 shows that for real substation data, the algorithm perform very well. In some sense, M_Prim may be comparable to well-known Simplex algorithm, which although known to be of exponential complexity, performs very well in practice with real data sets. Therefore, for larger networks where the ILP may fail to find an optimal solution, the M_Prim algorithm can be used as a substitute.

The work of Chapter 6 is further extended in Chapter 7 to study a more realistic network design where only a single CC is not considered. The Lagrangian Relaxation based approach proposed in Chapter 7, is another useful tool that can be used for larger data sets where the ILP will fail but very accurate upper and lower bounds for the solution can be

obtained using this approach and the heuristic approach can provide a near-optimal solution. One major drawback of the problem studied in Chapter 7 is that it did not take into account the load balancing which is the main goal of a distributed system. Incorporating proper load balancing among the LCCs while designing the optimal cost communication network can be addressed in future research and together with this, the optimal number of LCC placement can also be adopted. Redundancy in communication links is another important feature of smart grid ICT systems which is ignored in this dissertation for the purpose of simplification. However, in order to evaluate the exact cost of communication network design introducing optimal amount of redundancy among communication links can also be studied in the future.

Finally, in Chapter 8 this dissertation aimed at simultaneously studying the optimal PMU placement problem with minimum cost ICT network design and this study can be further extended by considering the assumptions in Chapter 7, with multiple LCCs. A heuristic solution for the problem studied in Chapter 8 can also be proposed in further studies which can guarantee a solution for the problem for very large data sets. Moreover, in Chapter 6, 7 and 8, only communication links with high bandwidth is taken into account. However, in reality ICT channels can have limited bandwidth and for that queuing delay should also be considered while formulating the solutions to the problems addressed in those chapters.

REFERENCES

[1] V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, and R. Setola, "Modelling interdependent infrastructures using interacting dynamical models," *Int. J. Critical Infrastructures*, vol. 4, no. 1-2, pp. 63-79, Dec. 2008.

[2] B. Falahati, and Y. Fu, "Reliability assessment of smart grids considering indirect cyber-power interdependencies," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1677-1685, Jun. 2014.

[3] M. Padhee, J. Banerjee, K. Basu, S. Roy, A. Pal and A. Sen, "A new model to analyze power system dependencies," 2018 *IEEE Texas Power and Energy Conference (TPEC)*, College Station, TX, pp. 1-6, 2018.

[4] J. Sanchez, R. Caire, and N. Hadjsaid, "ICT and electric power systems interdependencies modeling," in *Proc. Int. ETG-Congress Symp. 1: Security Critical Infrastructures Today*, Nov. 2013.

[5] A. Sen, A. Mazumder, J. Banerjee, A. Das, and R. Compton, "Identification of k most vulnerable nodes in multi-layered network using a new model of interdependency," in *NETSCICOM, INFOCOM WKSHP*. IEEE, 2014.

[6] S. Roy, H. Chandrasekaran, A. Pal and A. Sen, "A New Model to Analyze Power and Communication System Intra-and-Inter Dependencies", *2020 IEEE Conf. on Tech. for Sustainability (SUSTECH 2020)*, Santa Ana, pp.181-188, Apr. 2020.

[7] J. Banerjee, A. Das, C. Zhou, A. Mazumder and A. Sen, "On the Entity Hardening Problem in multi-layered interdependent networks," *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Hong Kong, pp. 648-653, 2015.

[8] K. Basu, M. Padhee, S. Roy, A. Pal, A. Sen, M. Rhodes and B. Keel, "Health Monitoring of Critical Power System Equipments using Identifying Codes", Luiijf, E., Žutautaitė, I., Hämmerli, B. (eds) *Critical Information Infrastructures Security. CRITIS 2018*. Lecture Notes in Computer Science, vol 11260. Springer, Cham, 2018.

[9] F. Ye and A. Bose, "Multiple communication topologies for pmu-based applications: Introduction, analysis and simulation," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5051–5061, 2020.

[10] X. Zhu, M. H. F. Wen, V. O. K. Li and K. Leung, "Optimal PMU-Communication Link Placement for Smart Grid Wide-Area Measurement Systems," in *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4446-4456, July 2019.

[11] M. Beg Mohammadi, R. Hooshmand and F. Haghighatdar Fesharaki, "A New Approach for Optimal Placement of PMUs and Their Required Communication Infrastructure in Order to Minimize the Cost of the WAMS," in *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 84-93, Jan. 2016,

[12] A. Pal, A. K. S. Vullikanti and S. S. Ravi, "A PMU Placement Scheme Considering Realistic Costs and Modern Trends in Relaying," in *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 552-561, Jan. 2017.

[13] A. J. Wood and B. F. Wollenberg, "Power generation, operation, and control", *John Wiley & Sons*, 2012.

[14] J. Wäfler, P.E. Heegaard, "Interdependency Modeling in Smart Grid and the Influence of ICT on Dependability.", In: Bauschert, T. (eds) Advances in Communication Networking. EUNICE 2013. Lecture Notes in Computer Science, vol 8115. Springer, Berlin, Heidelberg, 2013.

[15] X. Liu, B. Zhang, B. Chen, A. Aved and D. Jin, "Towards Optimal and Executable Distribution Grid Restoration Planning with a Fine-Grained Power-Communication Interdependency Model," in *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 1911-1922, May 2022.

[16] "Distribution automation feeder automation design guide," Cisco, San Jose, CA, USA, Rep. AN-1037-001_F, Aug. 2018. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/solutions/Verticals/Distributed-Automation/Feeder-Automation/DG/DA-FA-DG/DA-FA-DG.html.

[17] P. Chopade and M. Bikdash, "Critical infrastructure interdependency modeling: Using graph models to assess the vulnerability of smart power grid and SCADA networks," *2011 8th International Conference & Expo on Emerging Technologies for a Smarter World*, pp. 1-6, 2011.

[18] I. Shmulevich, E. R. Dougherty and Wei Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," in *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778-1792, Nov. 2002.

[19] A. Pal, C. Mishra, A. K. S Vullikanti and S. Ravi, "General optimal substation cover-age algorithm for phasor measurement unit placement in practical systems", IET Gener., Transm. Distrib vol. 11, no. 2, 347-353, Jan. 2017.

[20] A. Hamad and M. Kadoch, "SONET over Ethernet", in *Proc. 15 Annual Global Online Conference on Information and Computer Technology (GOCICT)*, Louisville, KY, USA, pp. 75-79, 4-6 Nov. 2015.

[21] D. Laidevant, F. Rambach and M. Hoffmann, “Availability of Connections in Ethernet over DWDM Core Networks”, in *Proc. 9th International Conference on Transparent Optical Networks*, Rome, Italy, pp. 24-27,1-5 July 2007.

[22] A. Comsa, A. B. Rus and V. Dobrota, "Simulation of the Floyd-Warshall algorithm using OMNeT++ 4.1," *2012 9th International Conference on Communications (COMM)*, Bucharest, pp. 229-232, May 2012.

[23] A. Abur, A.G. Exposito, *Power System State Estimation: Theory and Implementation*. New York: Marcel Dekken, 2004.

[24] M. Zhou, V. Centeno, J. Thorp, and A. Phadke, “An alternative for including phasor measurements in state estimators,” *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930-1937, Oct. 2006.

[25] H. Lin, S. S. Veda, S. S. Shukla, L. Mili and J. Thorp, "GECO: Global Event-Driven Co-Simulation Framework for Interconnected Power System and Communication Network," in *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1444-1456, Sept. 2012.

[26] J. Cui, A. Che, S. Li, & Y. Cheng, “Evaluation method on seismic risk of substation in strong earthquake area”, *PloS one*, *16*(12), Dec 2 (2021).

[27] Amir Bashian, Mohsen Assili, Amjad Anvari-Moghaddam, Omid Reza Marouz, “Co-optimal PMU and communication system placement using hybrid wireless sensors”, *Sustainable Energy, Grids and Networks*, 19, pp.100-238, 2019.

[28] S. Roy, “SSGMT: A Secure Smart Grid Monitoring Technique.”, In: Rashid A., Popov P. (eds) *Critical Information Infrastructures Security. CRITIS 2020*. Lecture Notes in Computer Science, vol 12332. Springer, Cham, 2020.

[29] S. Roy, “Secure Cluster Based Routing Scheme (SCBRS) for Wireless Sensor Networks”, In: Abawajy, J., Mukherjea, S., Thampi, S., Ruiz-Martínez, A. (eds) *Security in Computing and Communications. SSCC 2015*. Communications in Computer and Information Science, vol 536. Springer, Cham, 2015.

[30] J. Liang, L. Sankar and O. Kosut, "Vulnerability Analysis and Consequences of False Data Injection Attack on Power System State Estimation," in *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864-3872, Sept. 2016.

[31] E. Y. Song, G. J. FitzPatrick and K. B. Lee, "Smart Sensors and Standard-Based Interoperability in Smart Grids," in *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7723-7730, Dec. 2017.

[32] A. Roy, J. Bera and G. Sarkar, “Wireless sensing of substation parameters for remote monitoring and analysis”, in *Ain Shams Engineering Journal,* Vol.6, no. 6, pp. 95-106, 2015.

[33] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," in *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications*, pp. 113-127, 2003.

[34] G. S. Dhunna and I. Al-Anbagi, "A Low Power WSNs Attack Detection and Isolation Mechanism for Critical Smart Grid Applications," in *IEEE Sensors Journal*, vol. 19, no. 13, pp. 5315-5324, July 1, 2019.

[35] B. Sreevidya and M. Rajesh, "False Data Injection Prevention in Wireless Sensor Networks using Node-level Trust Value Computation," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2107-2112, 2018.

[36] H. F. Salama, D. S. Reeves, and Y. Viniotis, “The delay-constrained minimum spanning tree problem,” in Proceedings Second IEEE Symposium on Computer and Communications. IEEE, pp. 699–703, 1997.

[37] G. Dahl, L. Gouveia, and C. Requejo, “On formulations and methods for the hop-constrained minimum spanning tree problem,” in Handbook of optimization in telecommunications. Springer, pp. 493–515, 2006.

[38] G. Xue, “Minimum-cost qos multicast and unicast routing in communication networks,” IEEE Transactions on Communications, vol. 51, no. 5, pp. 817–824, 2003.

[39] J.M. Ho and D. Lee, “Bounded diameter minimum spanning trees and related problems,” in Proceedings of the fifth annual symposium on Computational geometry, pp. 276–282, 1989.

[40] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to algorithms. MIT press, 2009.

[41] “Haversine formula,” https://tinyurl.com/efpwtzak, accessed: 21-04-15.

[42] J. A. Bondy, U. S. R. Murty et al., Graph theory with applications.Macmillan London, 1976, vol. 290.

[43] [online] “Electric substations - hifld,” https://tinyurl.com/uc4466h4, accessed: 21-04-15.

[44] T. L. Baldwin, L. Mili, M. B. Boisen, and R. Adapa, “Power system observability with minimal phasor measurement placement,” IEEE Transactions on Power systems, vol. 8, no. 2, pp. 707–715, 1993.

[45] A. Sen, S. Roy, K. Basu, S. Adeniye, S. Choudhuri, and A. Pal, “Optimal cost network design for bounded delay data transfer from pmu to control center,” in 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, pp. 1–6, 2021.

[46] M. B. Mohammadi, R.-A. Hooshmand, and F. H. Fesharaki, “A new approach for optimal placement of pmus and their required communication infrastructure in order to minimize the cost of the wams,” IEEE Transactions on smart grid, vol. 7, no. 1, pp. 84–93, 2015.

[47] L. Gouveia, A. Paias, and D. Sharma, “Modeling and solving the rooted distance-constrained minimum spanning tree problem,” Computers & Operations Research, vol. 35, no. 2, pp. 600–613, 2008.

[48] [online] https://www.costowl.com/b2b/cabling-wiring/cabling-fiber-optic-cabling-cost/