Abstract

State estimation is the heart of many tools used in operations of distribution and transmission power systems. The quality of distribution systems state estimation (DSSE) typically suffers from a lack of adequate/accurate measurements and has not been fully implemented by many utilities. Recently, as part of many smart grid related initiatives to modernize power systems, electric utilities started to invest in advanced metering infrastructure (AMI) throughout their distribution systems. The main challenge in this area is that AMI measurements are generally not synchronized, and the difference between the measuring times of smart meters can be significant. In generation and transmission systems, the transmission system state estimation (TSSE) is already prevalent in many utilities. However, TSSE typically suffers from four major problems: partial unobservability, numerical ill-conditioning, bad data, and low accuracy. This thesis is based on three contributions. Firstly, an innovative method is developed to incorporate the non-synchronized measurements coming from AMI based on the credibility of each available measurement and appropriately adjusting the statistical property of the measurement signals. To illustrate the effectiveness of the proposed method, it is compared with traditional approach used in DSSE and the results show the improvements in the accuracy of DSSE. Next, based on the interconnection of the transmission system and distribution systems at PQ buses (feeder heads), a novel approach in TSSE method is presented which uses the DSSE results to provide additional measurements at the PQ buses of the transmission system. Comparisons between the traditional TSSE and the proposed TSSE show that significant improvements are achieved. The third contribution is the
methodology for identification of electricity theft points in distribution systems without violating privacy of consumers. The proposed approach models theft as bad data and consists of two stages. Firstly, the multiple bad data identification problem is solved using a heuristic optimization method to locate the points of theft which have redundant measurements. In the second stage, regarding identification of theft points which do not have redundant measurements, a method is proposed based on the discrepancies between the measured and estimated voltage magnitudes. Simulations results demonstrate the effectiveness of the proposed approach.
Preface

Based on the research presented in this thesis, several papers have been published in conference proceedings, and published and/or submitted as journal articles. In all publications, I have developed mathematical formulations, implemented models, conducted simulations, analyzed results, and prepared the initial manuscripts. My supervisors, Dr. Jatskevich and Dr. Vaahedi, have provided me with instructive comments and corrections throughout the process of conducting research studies, preparing and editing manuscripts. The following describes published and submitted papers as well as contributions of other co-authors.

Part of Chapter 2 was presented at a conference. A. Alimardani, S. Zadkhast, J. Jatskevich, E. Vaahedi, "Using smart meters in state estimation of distribution networks," PES General Meeting / Conference & Exposition, 2014 IEEE, vol., no., pp.1,5, 27-31 July 2014. As a colleague graduate student in our group, S. Zadkhast was working with me on this manuscript and provided comments and suggestions on the studies and content of this paper.

Chapter 2 has been published as a journal article. A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, E. Vaahedi, “Distribution System State Estimation Based on Non-Synchronized Smart Meters,” IEEE Transaction on Smart Grids, DOI: 10.1109/TSG.2015.2429640. I took advice from F. Therrien based on his experience from working for CYME International T&D on development of distribution system simulations. Also, Dr. Atanackovic, as the current manager of Real-Time System department, Grid
Operations of BC Hydro, BC., Canada, helped with explaining the asynchronicity problem of smart meters from industrial perspective.

Part of Chapter 3 was presented at a conference. A. Alimardani, S. Zadkhast, F. Therrien, J. Jatskevich, E. Vaahedi, "Impact of employing state estimation of distribution networks on state estimation of transmission networks," *PES General Meeting / Conference & Exposition, 2014 IEEE*, vol., no., pp.1,5, 27-31 July 2014. As colleague graduate students in our group, S. Zadkhast and F. Therrian worked with me on ideas on presentation of this manuscript and provided comments and suggestions on the studies and content of this paper.

Chapter 3 is submitted for peer review. A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, E. Vaahedi, “Incorporating Distribution System State Estimation Results at Transmission Level,” F. Therrien advised on the simulation process, and D. Atanackovic explained the industrial implementation of state estimation in DMS and EMS in BC Hydro, BC., Canada.

Chapter 4 is under preparation as A. Alimardani, J. Jatskevich, E. Vaahedi, “Identification of Electricity Theft in Distribution Systems”, to be submitted.
# Table of Contents

ABSTRACT .................................................................................................................................................. ii

PREFACE ................................................................................................................................................ iv

TABLE OF CONTENTS ........................................................................................................................ vi

LIST OF TABLES ....................................................................................................................................... x

LIST OF FIGURES ................................................................................................................................... xii

LIST OF ABBREVIATIONS ..................................................................................................................... xvi

ACKNOWLEDGEMENTS ........................................................................................................................ xvii

DEDICATION ........................................................................................................................................ xix

CHAPTER 1: INTRODUCTION .................................................................................................................... 1

1.1 Motivation .......................................................................................................................................... 1

1.2 Background of State Estimation ......................................................................................................... 4

1.2.1 Transmission system state estimation ............................................................................................ 4

1.2.2 Distribution system state estimation .............................................................................................. 8

1.2.3 Electricity theft detection ................................................................................................................ 11

1.3 State-of-the-Art Research ............................................................................................................... 12

1.3.1 Distribution system state estimation ........................................................................................... 12
1.3.1.1 DSSE based on AMI measurements ..................................................13

1.3.2 Transmission system state estimation ..................................................14

1.3.2.1 Enhancing TSSE using the results of DSSE .................................15

1.3.3 Electricity theft detection .................................................................16

1.4 Research Objectives and Anticipated Impacts ........................................17

CHAPTER 2: DSSE BASED ON NON-SYNCHRONIZED AMI MEASUREMENTS .................................................................20

2.1 Weighted Least Squares Approach ..........................................................20

2.2 State Estimation Based on Non-synchronized Smart Meters ..................24

2.2.1 Load variation modeling .................................................................25

2.2.2 Implementation of load variation modeling technique .......................30

2.3 Simulation of Case Studies .....................................................................32

2.3.1 IEEE 13-bus distribution system .......................................................37

2.3.1.1 Impact of random noise ..............................................................40

2.3.1.2 Impact of load uncertainty ..........................................................42

2.3.2 IEEE 123-bus distribution system ......................................................43

CHAPTER 3: INCORPORATION OF DSSE INTO TRANSMISSION LEVEL ..........................46

3.1 Transmission System State Estimation ..................................................46

3.2 Incorporating DSSE Results into TSSE ................................................47

3.3 Areas of Impact in TSSE .......................................................................50
CHAPTER 5: SUMMARY OF CONTRIBUTIONS AND FUTURE WORKS............. 94

5.1 Conclusions and Contributions ............................................................ 94
5.2 Potential Impacts of Contributions ...................................................... 98
5.3 Future Work .......................................................................................... 100

REFERENCES............................................................................................... 102
List of Tables

Table 2–1 Normality Evaluation of Figure 2-2 Data ................................................................. 28

Table 2–2 Statistical Performance of DSSE Error for a Typical Load Profile in the IEEE 13- 
bus Distribution System ........................................................................................................ 41

Table 2–3 Statistical Performance of DSSE Error for a Distorted Load Profile in the IEEE 
13-bus Distribution System .................................................................................................. 43

Table 2–4 Statistical Performance of DSSE Error for a Typical Load Profile in the IEEE 123-
bus Distribution System ...................................................................................................... 45

Table 2–5 Statistical Performance of DSSE Error for a Distorted Load Profile in the IEEE 
123-bus Distribution System ................................................................................................ 45

Table 3–1 Statistical Results of Calculating the Condition Number of TSSE for 100 Runs. 63

Table 3–2 Number of Iterations Taken by the Traditional and Proposed TSSE Methods 
Over 2 Hour Interval ............................................................................................................. 66

Table 3–3 Normalized Residuals of TSSE of the IEEE 14-bus Transmission System for Bus 
10 before Bad Data Detection ............................................................................................... 70

Table 3–4 Total Error of TSSE for the IEEE 14-bus Transmission System with Bad data at 
Bus 10 and Its Elimination .................................................................................................. 71
Table 4–1 Percentage of the Unmeasured Consumption at Each Theft Point.......................... 86

Table 4–2 Results of Multiple Bad Data Identification Solution Achieved by Different Methods ..................................................................................................................................................................... 87

Table 4–3 Comparison between Measured and Estimated Voltage Magnitudes Discrepancies with $3\sigma$ .................................................................................................................................................................................. 91
List of Figures

Figure 1—1 Typical architecture of EMS. Measurements from RTUs are send to the SCADA system. TSSE uses these data and provides other applications with required information... 5

Figure 1—2 Typical architecture of DMS. Measurements from the distribution power system are transferred based on which DSSE is executed providing necessary information for other applications................................................................. 10

Figure 2—1 Historical active power consumption profile of a building at The University of British Columbia................................................................. 26

Figure 2—2 Power consumption variation from 2:00PM to 2:15PM based on the historical data depicted in Figure 2-1................................................................. 28

Figure 2—3 DSSE process addressing asynchronicity of measurements.............................. 31

Figure 2—4 A typical load profile for 24 hours in pu with respect to the peak value based on the historical data recorded at The University of British Columbia. ......................................................... 32

Figure 2—5 Pattern of the updating time of each smart meter measurements installed on the loads of the IEEE 13-bus distribution system................................................................. 34

Figure 2—6 Simulation process for the case studies................................................................. 36
Figure 2—7 Norm error of real power measurements: (a) ideal case - error due to smart meters imprecision only; (b) error only due to out-of-date measurements, and (c) error due to smart meters imprecision combined with error due to out-of-date measurements. 38

Figure 2—8 Accuracy comparison of the three considered methods: (a) DSSE error during one day; and (b) accumulated error of DSSE during one day for each method. 39

Figure 2—9 Total error of DSSE in 24 hours for 10000 runs. 41

Figure 2—10 Typical and distorted load profiles for 24 hours in pu. 42

Figure 2—11 Considered IEEE 123 distribution systems. 44

Figure 3—1 Equivalent circuit for a transmission line. 46

Figure 3—2 Simplified diagram depicting a transmission system connected to distribution systems via PQ buses (feeder heads). 48

Figure 3—3 Configuration of the first case study: (a) Modified IEEE 14-bus transmission system with allocated loads, generators, and measurement devices; (b) The feeder head of the IEEE 13-bus distribution system is connected to bus 9 of the IEEE 14-bus transmission system. 57

Figure 3—4 Typical 24 hour load profile normalized with respect to the peak value based on historical data recorded at The University of British Columbia [91]; (b) Recorded 2-hour fragment of load profile from 2pm to 4pm. 58
Figure 3–5 Procedure of calculating the results of Proposed and Traditional TSSE at each step of the simulation.

Figure 3–6 IEEE 14-bus transmission system with a faulty critical measurement device. The subsystem which has a faulty measurement device is depicted. Red measurement devices show the injection power from the rest of the system to the subsystem.

Figure 3–7 Number of iterations taken to achieve a converged TSSE solution for: (a) the IEEE 14-bus transmission system; and (b) the IEEE-57 bus transmission system.

Figure 3–8 Total error of TSSE evaluated over all buses at each time step and its average for the 2-hour study interval with smart meters measurements updating every 15 minutes for: (a) the IEEE 14-bus transmission system; and (b) the IEEE 57-bus transmission system.

Figure 3–9 Total error of TSSE evaluated over all buses at each time step and its average for the 2-hour study interval when the smart meters measurements are updated every 30 minutes for the IEEE 57-bus transmission system.

Figure 4–1 A simplified diagram depicting connection of smart meters for modeling electricity theft as bad data: (a) load 2 is partially bypassing its smart meter to commit electricity theft; and (b) smart meter of load 2 is malfunctioning.

Figure 4–2 Probability density function of the chi-square distribution with 8 degrees of freedom.
Figure 4—3 The optimization process using SFLA......................................................................................... 84

Figure 4—4 Considered IEEE 123 distribution systems depicting eight loads with electricity theft. .............................................................................................................................. 85
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td>Advanced Metering Infrastructure</td>
</tr>
<tr>
<td>DMS</td>
<td>Distribution Management System</td>
</tr>
<tr>
<td>DSSE</td>
<td>Distribution System State Estimation</td>
</tr>
<tr>
<td>TSSE</td>
<td>Transmission System State Estimation</td>
</tr>
<tr>
<td>EA</td>
<td>Evolutionary Algorithms</td>
</tr>
<tr>
<td>EMS</td>
<td>Energy Management Systems</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>IPM</td>
<td>Interior Point Method</td>
</tr>
<tr>
<td>MA</td>
<td>Memetic Algorithms</td>
</tr>
<tr>
<td>MDMS</td>
<td>Meter Data Management Systems</td>
</tr>
<tr>
<td>NLP</td>
<td>Nonlinear Programming</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>OD</td>
<td>Out-of-date</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PSO</td>
<td>Particle Swarm Optimization</td>
</tr>
<tr>
<td>QP</td>
<td>Quadratic Programming</td>
</tr>
<tr>
<td>RTU</td>
<td>Remote Terminal Units</td>
</tr>
<tr>
<td>SA</td>
<td>Simulated Annealing</td>
</tr>
</tbody>
</table>
Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Juri Jatskevich and Dr. Ebrahim Vaahedi for their continuous support, patience, great inspiration, immense knowledge and insights, and enthusiasm that they provided me throughout my studies. I appreciate all their contribution of time and ideas to make my doctoral experience productive and stimulating. The financial support for this research was made possible by the Natural Science and Engineering Research Council (NSERC) of Canada under the Strategic Project Grant “Enabling Solutions for the Future Canadian Smart Grid” led by D. Jatskevich as a principal investigator.

I am also grateful to Dr. Atanackovic, the manager of the department of Real-Time Systems, BC Hydro, BC, Canada and a member of IEEE Working Group on State Estimation Algorithms, for his advice on a variety of issues regarding real-time power systems operations and practical concerns from the industry point of view as well as verifying simulations.

My sincere thanks also go to Prof. Uri Ascher. His lectures on numerical conditioning of systems were very helpful in my research. Also, I would like to thank Dr. Ebrahim Vaahedi and Mr. Nathan Ozog for the wonderful teaching experience I gained while assisting them.

In addition, I would like to thank my fellow graduate students and colleagues in Electric Power and Energy Systems research group at the University of British Columbia: Francis Therrien, Pouya Zadkhast, Arash Taavighi, Mehrdad Chapariha, and Yingwei Huang for
many stimulating discussions and all the memorable moments we have had in the past four years at UBC. Also, thanks to Eva Yu for proofreading parts of the text.

Lastly, I’d like to thank my parents. They supported me all along the way and I am grateful for everything they did for me.
Dedication

To My Wonderful Parents and Beloved Friends
CHAPTER 1: INTRODUCTION

1.1 Motivation

Modern electrical power and energy systems are fast evolving and undergoing significant changes to accommodate the new and renewable generation and satisfy the increasing power demand, while making the system more reliable, more resilient, and “smarter”. Significant resources are being deployed to enable the future smart grids [1], [2]. An integral part of the philosophy of smart grids is to use modern communication and advance metering technologies in power systems. As an investment towards Advanced Metering Infrastructure (AMI) development, the number of installed measurement devices in distribution systems has increased significantly in recent years [3]. Smart meters have been installed in very large numbers in many regions. For instance, in British Columbia (BC), Canada, BC Hydro is installing smart meters for every consumer throughout the province, which will provide 1.8 million accurate and reliable measurements [4]. Future distribution management systems (DMSs) are expected to be capable of handling a growing number of applications, including real-time control and monitoring, distribution transformer usage optimization, feeder reconfiguration and restoration, control of switches and reclosers, demand side management, and capacitor switching [5]-[10]. The deployment of distributed generation and renewable energy sources, which is crucial for enabling the
future smart grid, also poses new challenges, such as the occurrence of overvoltages at the
distribution level. Therefore, smart distribution systems are also expected to have volt/VAr
control capabilities. Since distribution system state estimation (DSSE) provides the initial
condition of all of the mentioned DMS applications, its accuracy and reliability will have a
significant impact on the operation of the future grid. Another recent trend in distribution
systems is the introduction of meshed topologies, which are incompatible with traditional
DSSE algorithms (or makes them less efficient) [1]. New DSSE techniques should be
capable of dealing efficiently with highly meshed and radial topologies.

A major challenge when using data from smart meters is that they do not provide
synchronized measurements (i.e. the measurement signals are not updated at the same
time) [11], [12]. Additionally, their measuring resolution is lower than measurements from
RTUs of reclosers. This problem has made utilities reluctant to use smart meters for some
real-time applications including state estimation. For instance, in many utilities such as BC
Hydro, smart meters are only considered as a tool for hourly billing and dynamic pricing.
Solving the problem with non-synchronized smart meter measurements is a necessity and
would significantly improve the capabilities of DMS with respect to distribution system
applications which are based on DSSE results. This objective is one of the motivations in
defining this thesis.

In energy management system (EMS), referring to transmission and generation
systems management, state estimation is in the heart of all major applications such as
transient stability analysis, voltage stability analysis, optimal dispatch, optimal reactive
power flow, preventing overloading, etc. [5]-[10]. One of the major problems with respect to TSSE is lack of sufficient measurement devices in lower voltage level parts of the system. For instance, the 25kV and 12kV substations in BC, Canada have this problem of inadequate measurement devices. As a result, the transmission system may be partially unobservable. Utilities typically overcome this problem by defining pseudo measurements to render the system observable. However, since pseudo measurements are assumptions based on historic data and tend to be not very accurate, their use also reduces the accuracy of state estimation. Another problem in TSSE is that the system is often numerically ill-conditioned. Numerous studies have been conducted to investigate numerical robustness of TSSE [13]. It is shown that increasing the redundancy of voltage magnitude measurements generally improves the numerical condition of the problem by reducing the condition number of system matrices [14]. However, installing Remote Terminal Units (RTUs) with measurement devices has very significant costs in transmission systems. Alternatively, if an AMI-based DSSE method with acceptable accuracy is developed, the results at feeder heads of distribution systems could be used to provide measurements for the TSSE PQ buses. These additional measurements could render the unobservable parts of transmission system to become observable, and at the same time improve numerical condition and increase the accuracy of TSSE without more investments on new measuring devices. Improvement in numerical condition of the system would further improve convergence and accuracy of TSSE. Bad data as a result of telemetry problems, measuring malfunction, or malicious attacks are a concern in TSSE. Since DSSE results at feeder heads provide more redundancy for measurements, bad data detection and identification in TSSE could be improved consequently.
The problem with AMI is that rich information exchange and hierarchical network structure increases the vulnerability for cyber-attacks [15], [16]. In particular, energy theft is a growing concern in both developing and developed countries [17]. Many studies have been conducted to address this problem based on load profile analysis. Analyzing the behavior of a load requires access to its consumption periodically, which is also recognized as private information. Disclosing such information could be a violation of privacy of customers [16]. This means that the electricity theft and privacy have become two contradictory objectives in distribution system management. Developing a method which can mitigate the electricity theft without disclosing load profile could incentivize utilities to continue investing in AMI and prevent a loss of billions of dollars in their revenue.

1.2 Background of State Estimation

The first step in power system operation is to monitor the current state of the system. This application is referred to as the state estimation (SE) [6], [13]. If the state of the system can be calculated based on the given measurements, the system is referred to as observable [18]. SE of transmission and distribution systems is impacted by the architecture of data acquisition from measurements devices in EMS and DMS.

1.2.1 Transmission system state estimation

Based on the estimated system state, other real-time system applications such as security analysis, automatic generation control, and optimal power flow with contingencies constraints, etc. can be executed. Figure 1-1 schematically depicts how the TSSE program is related to other real-time applications in a power systems control centre. As shown in
Figure 1—1 Typical architecture of EMS. Measurements from RTUs are send to the SCADA system. TSSE uses these data and provides other applications with required information.

Figure 1-1, in transmission systems, RTUs collect various types of measurements from the transmission power system and transfer them via supervisory control and data acquisition.
(SCADA) communication system. The SCADA system provides the TSSE program with the measurement data. The TSSE program consists of three tasks including

1) Observability analysis: Determining whether the system is observable based on the type and placement of measurements and the given system topology.

2) State estimation: Generally, measurements include line active and reactive power flows, line current magnitudes, bus voltage magnitude, generator outputs, loads, status information of circuit breakers and switches, transformer tap positions, and information of connected switchable capacitor banks. These measurements may not always be reliable due to the error of measurements, telemetry failures, and communication noise. These raw data are processed by the state estimator in order to filter the measurement noise and detect gross errors.

3) Bad data detection and identification: It may occur that the error of some measurements is significantly more than typical noise error. It is important to find a plausible explanation for the inconsistency in the data. Bad data may be caused by measuring devices' malfunctions and telemetry problems in the communication network. Several studies have been done in this field to detect and identify bad data [19]-[24]. This task identifies bad data and elimination of them from the state estimation procedure.

The results of TSSE program are used to build the network model. This model includes the system topology taking into account the status of switching devices and out of service
equipment, transmission line power flows, bus voltage magnitudes and angles, and
injection powers. Based on this model, different system applications are executed. These
applications can be categorized as real-time and offline applications. Real-time applications
contribute to real-time operation of the system. As shown in Figure 1-1, these applications
generate control actions and send them to the transmission system [25]. Off-line
applications analyze and evaluate the performance of the system for planning and design
purposes. They also predict long term demand growth for capital investment decisions.

Since the introduction of state estimation in transmission networks by Schweppe in late
1960s [26]-[28], the TSSE has benefited from a broad range of advances and developments.
In early stages, as there were not enough measurement devices installed in the system and
it was partially unobservable, the concept of observable islands was developed. An
observable island is an island for which all branch flows can be calculated from the
available set of measurements independent of the values adopted for reference values. As
more measuring devices have been installed in the last 4 decades, a larger portion of
transmission systems have become observable. Presently, most of the transmission
systems are observable in higher voltage levels—e.g., 60kV and higher. The observability of
transmission systems in lower voltage levels is usually obtained using pseudo
measurements.

Another common problem in TSSE is ill-conditioning which leads to inaccuracy and
divergence. In ill-condition problems, small errors in the entries could lead to significant
error in the solution [29]. In state estimation methods, a weight is assigned to each
measurement to represent its credibility [13]. Widely different weighting factors and the representation of low impedance branches could result in Ill-conditioning [30]. Adding more measurement devices could improve the numerical stability [14]. However, since it is a costly solution, utilities prefer various mathematical methods that have been developed to improve numerical conditioning as a widely accepted practice. These methods reformulate TSSE in order to increase robustness and achieve convergence in trade off with more computation. In addition, numerical conditioning methods may include partitioning the matrices into smaller ones and solving them in an interconnected fashion [30], [31]. Recently, researches have been done on deployment of phasor measurement units for monitoring and outage response purposes [32]-[35].

1.2.2 Distribution system state estimation

Similar to EMS, the DMS applications depend on the results of the state estimation program of the distribution system. Traditional DSSE methods have been developed with emphasize on three phase radial or weakly meshed topologies. One major problem in industrial DSSE implementation is that distribution systems are not observable. Therefore, DSSE developments gave rise to more studies on methods to create pseudo measurements. Pseudo measurements are constructed from the historical, sample load profile, and customer billing information [36]. Although this method is inexpensive and therefore desirable, it is typically very inaccurate and unreliable.

AMI is a recent development in distribution systems. The AMI measurements complement the existing RTUs measurements. This infrastructure includes smart meters, reclosers’
measurements, and their communication system, meter data management systems (MDMS), and means to integrate the collected data into application programs [37]. Another recent development in distribution networks is the introduction of meshed topologies which impacts some of the methods used in DSSE [1]. In radial networks, it was possible to trace the power flow from loads to the feeder and estimate the voltages merely based on pseudo measurements at the user end side. However, as non-radial distribution networks are being designed, DSSE techniques which are only applicable for these specific topologies should be revised or dismissed [38].

Figure 1-2 depicts typical architecture of DMS with recently developed AMI. In this figure, measuring devices of reclosers and substations and smart meters are the main sources of data, and smart meters are the dominant one. The number of smart meters installed in a typical distribution system is in the order of tens of thousands. Since certain utilities operate several distribution systems, it is not uncommon for large utilities to have a great number of smart meters. For example, BC Hydro has installed about 1.8 million smart meters throughout BC, Canada. As shown in Figure 1-2, due to the limited available bandwidth, smart meters do not incessantly transmit data to the aggregators (and then from the aggregator to the control center). Since synchronizing all smart meters with each other is considered too costly, smart meters practically do not transfer measurement signals simultaneously [12]. For example, each available measurement in the control center is a telemetered signal that has been measured somewhere between 0 to 15 minutes ago. Figure 1-2 shows that the reclosers use RTUs to transfer their measurement signals every 4 seconds.
Figure 1–2 Typical architecture of DMS. Measurements from the distribution power system are transferred based on which DSSE is executed providing necessary information for other applications.

During normal operation, the DSSE is executed e.g., every 5 minutes. However, there are several irregular event triggers in power control systems, such as drastic variation of RTU
signals and switching, which modify the topology of the system. Whenever an event trigger occurs, the DSSE must be executed again. Therefore, DSSE will not always be executed at regular intervals.

The smart meters measuring resolution is lower than the rate of DSSE execution. Therefore, for an available signal which is, for example, measured 10 minutes before, the error could be substantial as the load may have considerably varied since it was measured. This error, if ignored, could decrease the accuracy of state DSSE and negatively impact other applications which are based on the estimated system state.

Figure 1-2 shows real-time and offline applications using the results of DSSE. Similar to EMS, real-time applications issue control actions and send them to the system. Offline applications are exemplified by capital investment in Figure 1-2.

1.2.3 Electricity theft detection

With the AMI, not only can people read the meter remotely, but also implement customized control and demand response methods [39]. However, rich information exchange and hierarchical network structure in AMI expands the attack surface for metering to public networks and promotes vulnerability for cyber-attacks [40], [41]. There are different kinds of attackers to violate the AMI. Attackers in AMI can be classified as four categories [16]; Curious attackers that are only interested in the activity of their neighbors, greedy attackers that want to crack the AMI in order to steal electricity, malicious eavesdroppers attackers that collect data for vicious purposes such as house breaking, and active attackers that aim to compromise power systems to launch large-scale terrorist attacks.
Among various violation of AMI, energy theft in particular is a growing field in both developing and developed countries [16]. A World Bank report shows that up to 50% of electricity in developing countries is supplied through theft [42]. Each year, over 6 billion dollars are lost in the United States as a result of energy theft [43]. In 2009, the FBI reported an organized energy theft attempt that may have cost up to 400 million dollars annually for a utility after installing AMI [16]. In Canada, BC Hydro reported a 100 million dollars of losses every year [44]. In India and Brazil, utility companies incur losses around $4.5 billion and $5 billion, respectively [16]. In Netherlands, non-technical loss represents about 23% of total loss with an additional loss of about 1200 GWh/year for illegal use [45]. After installing new meters at 5% of the points of delivery in Italy to check power usage of downstream customer meters, the theft detection increased up to 50%.

Since electricity theft has greatly impacted the revenue of utilities, some companies have started to invest on implementation of methods to counter this problem.

1.3 State-of-the-Art Research

1.3.1 Distribution system state estimation

Researchers have been working on enhancing the DSSE as many advance technologies are being deployed in distribution systems. Baran and Kelly proposed a DSSE technique for real-time monitoring of the distribution system based on the Weighted Least Squares (WLS) approach and uses a three-phase node voltage formulation [46]. Branch-current based DSSE was developed in [47] and [48]. Similarly, different methods were proposed to
solve the DSSE for unbalanced and/or asymmetric systems [47], [49], [50] and radial networks [51]-[53]. However, introduction of meshed topologies in future smart distribution networks renders some of these methods inapplicable [1]. Admittance matrix based methods is introduced in [54] and improved in [55]. Therrion et al. proposed a DSSE which unifies power flow and short-circuit calculation algorithms to achieve a unified DSSE with desirable numerical characteristics using modified augmented matrices [56].

1.3.1.1 DSSE based on AMI measurements

At this point, few studies focused on the DSSE based on smart meters can be found in the literature. In [57], a DSSE method based on synchronized AMI measurements is established. Studies in [58] show that smart meters can improve the accuracy of DSSE assuming that, similarly to transmission networks, the measurements of smart meters are reasonably synchronized. However, treating the signals as if they are all synchronized could reduce the quality of DSSE since the time difference between the measurements may be significant in practice. To take the asynchronicity of smart meters into account, [59] assumes available smart meter measurements have two levels of error (2% or 10%), while [11] considers all smart meters to have the same error (10%). However, the approaches in [59] and [11] do not take into account the historical short-term load variation patterns nor the length of time that has passed between updating of the measurement and the execution of DSSE. Alternative approach includes modeling pseudo loads using a neural network (NN) in for DSSE applications [60]. Despite the contributions to modeling pseudo measurements, the problem with this method is that the high uncertainty and poor accuracy of the generated pseudo measurements for real-time applications.
1.3.2 Transmission system state estimation

There have been numerous studies on TSSE over the last four decades. The computational burden of traditional TSSE is significantly reduced by fast decoupled TSSE in [31]. In order to analyze the observability of the system, topological and numerical analysis methods are developed [61]-[64]. Based on these methods, in order to ensure full network observability, an optimal measurement placement method is proposed in [18]. Various studies have tried to deal with the numerical ill-conditioning of TSSE. Measurements which are based on topological information, such as zero injection when there is no load at a bus or zero current for open circuits, are referred to as virtual measurements. Authors of [65] proposed a method that improved the numerical conditioning by formulating normal TSSE equations in combination with the virtual measurements that are considered as constraints. By separating the residuals as independent variables, [66] proposed a numerically stable method for TSSE, which is still practiced in state estimation of transmission and distributions systems. In [67], orthogonal transformation method is applied in order to improve the numerical condition of TSSE, which is widely accepted in industry. Later, it is shown that orthogonal method in hybrid with normal equations would significantly improve the numerical stability and solution accuracy compared against alternatives, such as the Peters-Wilkinson and Hachtel methods [68]. However, this method is computationally expensive. Normalized residuals were used in order to develop a basic bad data identification method in TSSE [24]. To evaluate the quality of TSSE results, hypothesis test is developed to check the solution and detect the existence of bad data [69].
Later, in order to identify multiple bad data in measurements, [70] proposed a bad data identification test as an optimization problem.

**1.3.2.1 Enhancing TSSE using the results of DSSE**

The possibility of using DSSE for enhancing TSSE has yet been fully exploited. Historically, utilities developed energy management systems and distribution management systems separately, and TSSE and DSSE have often been pursued by independent departments. Moreover, until recently, the results of DSSE were unavailable (due to unobservability) or too inaccurate to be considered useful. Consequently, incorporating these results into TSSE was unappealing. To the best of the authors’ knowledge, only a few studies on relating/combining TSSE and DSSE are available in the literature. In [71], a SE algorithm is proposed for observable transmission systems and unobservable distribution systems. The method is implemented in two stages. Typical SE is first executed at the transmission level, followed by an external estimation of the unobservable distribution systems. Therefore, TSSE is not impacted by the DSSE results. Reference [72] proposes a global SE framework to simultaneously perform SE for transmission and distribution systems. Despite the attempt of proposing a comprehensive method, the primary problem with this approach is that, realistically, each of the distribution systems connected to the transmission system has thousands of buses, resulting in a very large overall system of equations. More importantly, the commonly used industrial SE programs would need to be modified to a great extent in order to allow iterations between DSSE and TSSE, which is not easily accomplished in practice.
1.3.3 Electricity theft detection

Several methods to mitigate energy theft have been proposed recently. The first category of these methods is based on classification techniques. The idea is to study the load profile over a period of time and extract patterns to distinguish abnormal energy usage form normal energy usage [73]-[83]. Machine learning and data mining technologies are used to generate a good classifier based on sample data. The problem is that in many practical situations, an example of the attack class is not obtainable [16]. Moreover, most of the methods in this category are based on the assumption that the attackers are not adaptive and do not try to evade the detection mechanisms, leading to missing intelligent energy theft. As a result, detection rate can be affected significantly [16]. The second category for electricity theft is state-based methods. In these methods, the goal is to detect theft by monitoring the state of the system. This monitoring could be provided by AMIs, wireless networks, mutual inspection, etc. in combination with modeling and statistical methods [15], [17]. The main problem with these methods is the cost of extra investment by utility companies, which includes measurement device costs, software costs, system implementation costs and training costs. Third category of methods for electricity theft is game theory based. This is a relatively recent technique for electricity theft. In the developed methods, a tariff and investment strategy is proposed from a game theory point of view; assuming that involving actors are rational. These methods are not mature yet, but they provide a new perspective to recognize electricity theft [84], [85].

Privacy preservation is one of the main challenges in practical implementation of the developed theft identification methods. Using information about the behavior of a load,
including load profiles, is recognized as a violation of privacy, and would raise concerns [16]. In particular, the private information of customers may be sold to third parties for marketing or to insurance companies. It could also be used for criminal activities; robbers could analyze the pattern of consumption by a potential victim to deduce their daily activities and behavior. The consumption pattern could also be used to determine household routines and vacancy times. Many researchers have called for legislators and authorities to address this threat [16]. This means that electricity theft and privacy have become two contradictory notions in distribution system management.

1.4 Research Objectives and Anticipated Impacts

Objective 1: DSSE based on AMI measurements

This first objective of this thesis addresses the asynchronicity problem of AMI measurements in DSSE. The proposed method should improve the accuracy of DSSE in comparison to the traditional method which treats the non-synchronized measurements as synchronized. Comparison of traditional and the proposed DSSE on different case studies should be delivered. Since the measurement noise is a random phenomenon, the performance of the proposed method should be tested different measurement noise, and it should consistently outperform the traditional DSSE. The observability analysis should be presented in order to explain whether or not the AMI measurements will result in an observable system.
The proposed algorithm should require minimum modifications to the existing DSSE algorithms in order to be appealing for practical implementation and usage in industrial applications and commercial packages. Moreover, the proposed method should not depend on any specific distribution system topology, and it should be effective for both radial and meshed distribution systems.

**Objective 2: Enhancing TSSE using the results of DSSE**

As the second objective, this thesis presents an innovative and practical framework that utilizes the estimated feeder heads power flows and bus voltage magnitudes as equivalent power injection and voltage magnitude measurements in transmission systems. Specifically, the feeder heads power flows and voltage magnitudes calculated by the DSSE algorithm (and their variances) need to be transmitted to the TSSE program. Moreover, the estimated voltage magnitude at distribution substation is used as the PQ bus voltage magnitude measurement. In the proposed approach, the DSSE and TSSE programs should may be executed independently using different commercial packages, making this approach practical for industry.

It is important to investigate and demonstrate how the DSSE results could restore the observability of the unobservable parts of transmission network and thus, enable the execution of TSSE. The proposed approach should also be examined regarding its impact on the numerical conditioning of TSSE and the number of iterations that TSSE takes to converge. Moreover, the impact of the new approach in facilitating bad data identification in transmission systems should be shown. Finally, the impact of AMI-based DSSE
measurements, better numerical conditioning, and higher rate of success in bad data identification on accuracy of TSSE should be presented.

**Objective 3: Electricity theft detection**

The goal of this objective is to enhance the DMS by identifying electricity theft points in the distribution system using the DSSE method developed in Objective 1. Electricity theft at a load is modeled as (hypothetical) malfunctioning of the measurement device with no theft at that load. Therefore, bad data identification methods will find the measurements which are either the results of truly malfunctioning measuring devices or the results of electricity theft. A criterion should be presented to distinguish as many cases of theft as possible from truly malfunctioning devices. Moreover, the problem with preserving the privacy of consumers should be considered, which implies that using the customer’s consumption profile is not desirable and should be avoided.
CHAPTER 2: DSSE BASED ON NON-SYNCHRONIZED AMI MEASUREMENTS

2.1 Weighted Least Squares Approach

Different state estimation techniques have been developed and used for transmission and distribution systems. The most established approach is the weighted least squares (WLS) method [13], [86], [87]. The formulation of WLS DSSE is based on the linearization of the relationship between measurements and state variables. The nonlinear relationship between the state vector and the measured electric variables can be stated as

\[ z = h(x) + v \]  \hspace{1cm} (2-1)

where \( z \) denotes the vector containing the measurements, \( x \) represents the state variables, vector function, \( h(x) \) relates the measurements to the state variables, and \( v \) represents the noise in the measurements. Several combinations of variables can be used to form \( x \). In this section, the state variables are defined as the angle and magnitude of voltages at all buses, for a total of six state variables per three-phase bus [81]. As an example, in a system with 3-phase buses, the active and reactive powers injected at bus \( i \) on phase \( p \) can be written as a function of state variables as:

\[ P_i^p = \sum_{l=1}^{n} \left[ V_{i}^p \left| G_{i,k}^{p,l} \cos(\theta_{i,k}^{p,l}) + B_{i,k}^{p,l} \sin(\theta_{i,k}^{p,l}) \right| \right] \]  \hspace{1cm} (2-2)
\[ Q_i^p = \left| V_i^p \right| \sum_{k=1}^{n} \sum_{l=1}^{1} V_{i,k}^p \left[ G_{i,k}^p \sin(\theta_{i,k}^p) - B_{i,k}^p \cos(\theta_{i,k}^p) \right] \]  

(2-3)

where \( n \) is the number of 3-phase buses, and \( Y = G + jB \) is the system admittance matrix.

Equations (2-2) and (2-3) are used to construct \( h(x) \).

A power system measurement can be considered as a random variable with Gaussian (normal) probability density function (PDF). The mean of this variable is the actual value of the variable which is measured, and the standard deviation of the random variable represents the error of the measurement. The Gaussian PDF of a measurement is defined as

\[ f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left( \frac{z-\mu}{\sigma} \right)^2} \]  

(2-4)

where \( \sigma \) is the standard deviation and \( \mu \) is the mean. This function is also referred to as the likelihood function [13]. For a system, the joint PDF (which represents the probability function of \( m \) independent measurements with Gaussian PDF) can be expressed as the product of individual PDFs (assuming that each measurement is independent of the rest) as [13]

\[ f(z) = f(z_1)f(z_2)\ldots f(z_m) \]  

(2-5)

where \( z_i \) is the \( i \)th measurement and \( z = [z_1, z_2, \ldots, z_m] \). For a given set of measurements and their corresponding standard deviation, the likelihood function (joint PDF) will attain its peak when the unknown mean values are chosen closest to their true values. Thus, the goal is to maximize the likelihood function by assigning mean values which are closest to
the measured values. The mean value for a measured electric variable is a function of state variables, which can be expressed as

\[ \mu_i = h_i(x_1, x_2, \ldots, x_n) = h_i(x). \]  

(2-6)

Based on the assumption that the mean of the measurement noise is zero, the explained process is formulated as

\[
\max \ f(z) \equiv \max \ \log(f(z)) = -\frac{1}{2} \sum_{i=1}^{m} \left( \frac{z_i - \mu_i}{\sigma_i} \right)^2 - \frac{m}{2} \log 2\pi - \sum_{i=1}^{m} \log \sigma_i.
\]  

(2-7)

In (2-7), without loss of generality, the log function is used to simplify the equations. Since the last two terms in (2-7) are constant, maximizing \( f(z) \) is equivalent to

\[
\min \ J(x) = \sum_{i=1}^{m} \left( \frac{z_i - \mu_i}{\sigma_i} \right)^2
\]  

(2-8)

Accordingly, the objective function of the minimization problem is rewritten as

\[
J(x) = \sum_{i=1}^{m} \left( \frac{z_i - \mu_i}{\sigma_i} \right)^2 = \sum_{i=1}^{m} \left( \frac{z_i - h_i(x)}{R_{ii}} \right)^2 = [z - h(x)]^T R^{-1} [z - h(x)]
\]  

(2-9)

where \( R = \text{diag} \{ \sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2 \} \).

This means that the inverse of the variance weigh each measurement, i.e. measurements with larger variance have a lesser impact on the optimization process.
The optimality condition is satisfied at the minimum point of $J(x)$. Therefore,

$$g(x) = \frac{\partial J(x)}{\partial x} = -H^T(x)R^{-1}[z - h(x)] = 0$$  \hspace{1cm} (2-10)

where $H(x) = \left[ \frac{\partial h(x)}{\partial x} \right]$ is the measurement Jacobian.

To solve (2-10), the Gauss-Newton method is employed which iteratively calculates $x$ using

$$x^{k+1} = x^k - [G(x^k)]^{-1}g(x^k)$$  \hspace{1cm} (2-11)

where $G(x) = \frac{\partial g(x)}{\partial x} = H^T(x)R^{-1}H(x)$ is referred to as the gain matrix, and the superscript $k$ is the iteration number. Various considerations such as equality constraints for zero injection points could be added to the equations above as explained in [81].

Generally, the vector $z$ includes the measurements at different substations, buses, and branches, i.e., bus voltage magnitudes, active and reactive power injections, transmitted active and reactive powers, and line currents. These measurements should be sufficient to ensure that there is a unique solution of $x$. If there are enough measurements to uniquely determine $x$, the network is referred to as observable.

In traditional distribution systems, the network is typically unobservable due to an insufficient number of measurements. Therefore, inaccurate and/or unreliable pseudo measurements are often employed in DSSE making its solution also inaccurate and/or
unreliable. In smart distribution systems, active and reactive powers of all the PQ buses are measured by smart meters. In addition, voltage angle is typically assigned at the feeder head, which is analogous to the slack bus in transmission networks. Hence, similarly to the case of power flow problem in transmission systems with known injection powers, smart distribution networks can be considered as observable. Distribution systems commonly contain voltage magnitude and power measurements in substations and a number of reclosers along the feeders. Those measurements provide some redundancy to the system measurements.

### 2.2 State Estimation Based on Non-synchronized Smart Meters

Typically, smart meter are assumed to update their measured electric variables signal every 15 minutes [88], [89]. These measurements, which are not updated during this interval, are referred to as out-of-date (OD) signals. Referring to 0(2-9), the credibility of a measurement in WLS-based DSSE is represented by the inverse of its variance. Since consumer loads change constantly, OD signals are expected to have additional error. Specifically, if a smart meter has measured the power and voltage of a load some minutes ago, the total error of the available measurement signal can be considered to come from two sources: i) the smart meter’s own intrinsic imprecision; and ii) the result of variation of load consumption since it was measured. As the measurement sample is updated every 15 minutes, until the new sample signal is transferred, the available OD signal will not track the changes of the load within the 15 minute interval. In the proposed method, the error caused by the short-term load variation is modeled as a random variable. In the following, the properties of this random variable are studied.
2.2.1 Load variation modeling

The error of an OD signal at any point is thus assumed to consist of two components: the error of the measurement device; and the error resulting from the short-term load variation. Therefore, the total error $e_{\text{total},i}$ of the $i$th meter could be represented as

$$e_{\text{total},i} = e_{m,i} + e_{OD,i}$$  \hspace{1cm} (2-12)

where $e_{m,i}$ is the device’s measurement error, and $e_{OD,i}$ is the load variation error. The objective of this section is to model the total error in a manner which is compatible with the WLS-based DSSE process presented in Section 2.1.

The first component of the total error has been well studied. It was concluded to roughly follow a Gaussian PDF with zero mean and a specific standard deviation [13]. Generally, the precision of a measuring device is given by the manufacturer as determined through several tests.

The second component of the total error is the variation of the load from the moment it was sampled to the time when DSSE is executed. In some studies, the consumption of residual loads has been studied, and it is suggested that the load can be modeled as normal, log-normal, and gamma distributions [90]. However, to the best of our knowledge, no study has been conducted to assess the statistical distribution of short-term distribution-level variation of load consumption over different points of a day.
To do so, the historical data from a sampled building consumption at The University of British Columbia campus has been analyzed [91]. Its active power consumption sampled every 15 minutes over more than 1 month is represented in Figure 2-1.

![Figure 2-1](image)

**Figure 2-1** Historical active power consumption profile of a building at The University of British Columbia.

2-1. The objective is to verify whether the variation of the load from any time of every day to the next sampling point follows normal distribution characteristic. If so, then the total error could be modeled as the combination of two normally distributed random variables which is also a normal variable [92] and could be incorporated in the DSSE process as

\[
\mu_{\text{total}} = \mu_m + \mu_{\text{OD}} = \mu_{\text{OD}}
\]  

(2-13)
\[ \sigma_{\text{total}}^2 = \sigma_{m_i}^2 + \sigma_{\text{OD}}^2 \]  \hspace{1cm} (2-14)

where \( \mu_{\text{total}} \), \( \mu_{m_i} \), and \( \mu_{\text{OD}} \) are the mean values of the total, measurement device, and OD signal errors; and \( \sigma_{\text{total}} \), \( \sigma_{m_i} \), and \( \sigma_{\text{OD}} \) are the standard deviations of the total, measurement device, and OD signal errors, respectively. Therefore, some tests are required to evaluate the normality of the load variation.

There are a handful of fitness tests to check whether or not a set of data is normally distributed [93]-[96]. Two of the most reliable tests are the Shapiro-Wilk and Anderson-Darling tests [95], [97], [98], which are implemented in this study. Presuming that a signal is normally distributed (this assumption is referred as the ‘null hypothesis’), each test attempts to calculate an indicator which is called the p-value to evaluate the normality of a signal. If the p-value is less than a threshold (0.05 based on [99]), the null hypothesis is refuted meaning that the signal is not normally distributed. To apply these tests in our study, we first calculate the load variation for any time of the day for several days. For instance, Figure 2-2 shows the signal of load variation (active power) from 2pm to 2:15pm for 38 days. Once these signals are prepared, they are fed to the tests to see whether they follow a normally distributed pattern. The results of the normality tests for the data in Figure 2-2 are presented in Table 2-1. As one can see, the p-value for both tests is clearly above 0.05. This means that the null hypothesis is not refuted and that it is reasonable to assume that the signal has a normal distribution characteristic. This table includes the same analysis for reactive power. The p-value for other points of the day are also found to be well above the 0.05 threshold, which means that the second component of the error in
Figure 2—2 Power consumption variation from 2:00PM to 2:15PM based on the historical data depicted in Figure 2-1.

Table 2—1 Normality Evaluation of Figure 2-2 Data

<table>
<thead>
<tr>
<th>Normality Evaluation Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis on Active Power</strong></td>
<td></td>
</tr>
<tr>
<td>Anderson-Darling Test</td>
<td>0.2282</td>
</tr>
<tr>
<td>Shapiro-Wilk Test</td>
<td>0.3367</td>
</tr>
<tr>
<td><strong>Statistical Analysis on Reactive Power</strong></td>
<td></td>
</tr>
<tr>
<td>Anderson-Darling Test</td>
<td>0.2495</td>
</tr>
<tr>
<td>Shapiro-Wilk Test</td>
<td>0.3218</td>
</tr>
</tbody>
</table>

(2-12) (i.e. resulting from the signal being OD) can be modeled as a random variable with normal distribution. Thereafter, the mean and standard deviation at any point of the day
and for any time distance (between 0 and 15 minutes for this case study) can be calculated accordingly. It is worth mentioning that the mean value of load variation as depicted in Figure 2-2 as an instance is very close to zero. This shows that in fact, it is mainly the adjustment of variance which is changing the results.

The load historical data of vast majority of utilities in distribution system is not available at any sampling rate higher than the resolution of smart meters. Without loss of generality, it can be assumed that the mean and standard deviation vary linearly in the interval between two consecutive events of updating measurements. For an OD signal, the time passed since the last update is represented by the variable \( t_{LU} \). If smart meters update their signals every 15 minutes, \( t_{LU} \) can be anywhere between 0 to 900 seconds. The variable \( t_{LU} \) is reset to 0 every time the signal is updated. As a result, the mean and standard deviation employed in the DSSE process are calculated as

\[
\mu_{OD}(t_{LU}) = \frac{t_{LU}}{t_{LU_{\max}}} \times (\mu_{OD_{max}}(t_{LU_{\max}}) - \mu_{OD_{0}}(0)) + \mu_{OD_{0}}(0) \\
\sigma^2_{OD}(t_{LU}) = \frac{t_{LU}}{t_{LU_{\max}}} \times (\sigma^2_{OD_{max}}(t_{LU_{\max}}) - \sigma^2_{OD_{0}}(0)) + \sigma^2_{OD_{0}}(0)
\]

(2-15)

(2-16)

where \( \mu_{OD_{0}}(0), \mu_{OD_{max}}(t_{LU_{max}}), \sigma_{OD_{0}}(0), \) and \( \sigma_{OD_{max}}(t_{LU_{max}}) \) are the calculated mean and standard deviation of load variation of the first and the last points of the interval, and \( t_{LU_{max}} \) is the interval between two consecutive updating events. Therefore, by providing the
mean and standard deviation at updating points, DSSE could be performed at any point within the interval.

### 2.2.2 Implementation of load variation modeling technique

To implement the proposed technique in existing DSSE algorithms, the mean and standard deviation of the load variation at each measurement updating point based on historical data are first calculated. Every time that the DSSE process is being executed, the variance of each measurement is updated by evaluating (2-14) and updating \( \sigma_i^2 \)s in the \( R \) matrix accordingly. As mentioned, in (2-7), it is assumed that the mean value of the noise measurement error is zero. However, in the proposed method, the mean value of the total error is calculated based on (2-13) and (2-15). Given that the measurement is modeled as

\[
z = h(x) + v + \Delta L
\]  

(2-17)

where \( \Delta L \) is the load variation vector as a result of OD measurements, the objective function in (2-7) is modified to

\[
\max f(z) = \max \log(f(z)) = \\
\frac{-1}{2} \sum_{i=1}^{m} \left( \frac{z_i - \mu_i - \mu_{OD}}{\sigma_i} \right)^2 - \frac{m}{2} \log 2\pi - \sum_{i=1}^{m} \log \sigma_i
\]  

(2-18)

where \( \mu_{OD} \) is the mean of load variation calculated in (2-15). Therefore, the mean of the error must be subtracted from the measured value for each measurement, and (2-10) would be modified to
where $\mu$ is a vector containing the means of all the load variation errors.

A flowchart summarizing the proposed approach is shown in Figure 2-3. It should be noted that the proposed method is general in the sense that it can be incorporated in all WLS-based DSSE algorithms, independently of their sets of state variables or their ability to model radial and/or meshed topologies.
It is important to mention that the DSSE is an iterative method which requires calculation of active and reactive powers from magnitude and angles of the voltages, calculating the Jacobian for the next iteration to solve the WLS optimization problem, and matrix based calculations (including factorization of the gain matrix) to obtain the solution. The proposed method adds a very simple set of operations (2-13)-(2-16), calculated only at the beginning of the DSSE process (not in each iteration), which is numerically insignificant in comparison to the overall DSSE iterative solution of the WLS problem as explained above.

2.3 Simulation of Case Studies

In this section, the effectiveness of the proposed method is examined using the IEEE 13-bus and 123-bus distribution test systems [100]. Each case study includes the execution of DSSE every 200 seconds over the course of 24 hours, during which the loads change following a pattern similar to Figure 2-4. Every 200 seconds, initially, a 3-phase unbalanced

![Figure 2-4 A typical load profile for 24 hours in pu with respect to the peak value based on the historical data recorded at The University of British Columbia.](image-url)
power flow is performed to obtain a reference solution in accordance to the new loads. Afterwards, Gaussian noise is added to the actual value of electric variables in order to emulate the measurements’ error. As mentioned, it is assumed that each smart meter’s measurement signals are updates every 900 seconds. However, smart meters are not synchronized. To simulate asynchronous measurements, a random integer between 0 and 900 is assigned to each smart meter. This number represents the time of first measurement update for each smart meter from the starting point of the simulation. At the starting point of the simulation, it is assumed that all the smart meter measurements are available. Afterwards, each smart meter’s measurement signal is updated at the assigned time to it. For example, the assigned times for the IEEE 13-bus system are depicted in Figure 2-5. Following the
first measurement, each smart meter’s measurement signal is updated every 900 seconds. In this interval, DSSE uses the last updated signal of the smart meter (OD measurement). The precision of power measurements is assumed to be 1% [101]. The uncertainty of network parameters is considered using a uniform distribution [102]. The considered bound of uncertainty is 5%.

For comparison purposes, three approaches are implemented. Figure 2-6 summarizes the simulation process. In the first approach, it is assumed that all the smart meters transferring data to the control center incessantly. This is an ideal assumption (which is not realistic), and it is only considered for comparison and providing a better insight. This
approach with ideal assumption will be referred to as “Ideal DSSE”. In the second and third approaches, the smart meters measurements are updated every 900 seconds following the approach explained in the previous paragraph. In the second approach, the traditional DSSE is performed and the asynchronicity of measurements is not considered, whereas in the third approach, the proposed method of adjusting the variance is employed. These methods are referred to as “Traditional DSSE” and “Proposed DSSE”, respectively. After having computed the state vector using each method, the error of DSSE can be calculated. In this study, the 2-norm error of DSSE at each execution is defined over all bus voltages as

$$error = \sqrt{\sum_{i=1}^{no\_bus} \left| (v_{i}^{est} - v_{i}^{act}) \right|^2}$$

(2-20)

where $no\_bus$ is the number of buses (including all single phase buses), $v_{i}^{est}$ is the estimated voltage of the $i$th bus, and $v_{i}^{act}$ is the actual (errorless) value of $i$th bus voltage. Each voltage variable in this equation is a complex number in per unit. The accumulative error in (2-20) over the 24 hour simulation interval is referred to as the total error of the estimation.
Figure 2—6 Simulation process for the case studies.

As Figure 2-6 shows, the proposed method has an additional block compared to the traditional DSSE which is explained in Section 2.2. The measurements of the traditional
DSSE and the proposed DSSE are identical whereas the measurements of the ideal DSSE are always up to date as explained.

2.3.1 IEEE 13-bus distribution system

The IEEE 13-bus distribution test system [100] is used for the first case study. This 3-phase system is unbalanced with 32 single-phase buses. For the considered system, the 2-norm of all errors in measurements of active power [calculated similarly to (2-20)] at each execution of DSSE for three cases are shown in Figure 2-7. In case (a), it is assumed that the smart meters measurement signals are always up-to-date and the measurement errors result only from the imprecision (i.e. noise) of the measuring devices. In case (b), the error due to OD measurements, which depends on the variation of the load since the last measurement, is shown. This portion of the error becomes especially pronounced during increases and decreases of the power demand, as can be observed in Figure 2-7. Case (c) is the combination of cases (a) and (b), i.e. the summation of the device imprecision and OD measurement errors.

To demonstrate the performance of the proposed method, the DSSE error of each method has been calculated using (2-20) and the results are shown in Figure 2-8. Specifically, the DSSE error achieved by each method during the day is depicted in Figure 2-8(a). To demonstrate the cumulative measure of these errors, the DSSE errors are also integrated and the results are shown in Figure 2-8(b). As expected, the best state estimation results
are obtained under the ideal case when all smart meters are assumed to be always up-to-date (see Figure 2-8, “Ideal DSSE”). Additionally, it is seen in Figure 2-8 that the proposed DSSE method results in a noticeable reduction of the total accumulated error over the day to 12.89 compared with to the traditional DSSE method (15.08). It is also observed that as the load changes during the day (see Figure 2-4), the DSSE error is increasing when the load varies greatly, i.e. during the morning load pickup and evening load decrease. However, when the rate of load variation is low, the estimation error is small, and the
Figure 2–8 Accuracy comparison of the three considered methods: (a) DSSE error during one day; and (b) accumulated error of DSSE during one day for each method.
results of the traditional DSSE and the proposed DSSE are almost the same, which can also be observed in Figure 2-8(a). As expected, the DSSE error reproduced in Figure 2-8 are in good agreement with the measurement errors depicted in Figure 2-7. While the traditional DSSE and proposed DSSE have the same error in their precision, the proposed method treats the measurements differently by appropriately adjusting their variances, and therefore achieves higher accuracy.

2.3.1.1 Impact of random noise

Since the measurement noise is assumed to be random, it can be expected that the results will change every time the simulation is run. To consider this issue, the simulation has been run 10,000 times, and the total accumulated over 24 hours DSSE error for each of the methods is depicted in Figure 2-9. Therein, it is observed that the random nature of noise has a negligible impact on the total error of each method. Additionally, the best, worst, and mean performances for each method are summarized in Table 2-2. As also observed in Figure 2-9 and Table 2-2, the impact of measurements noise is minimal, e.g. the total DSSE error of the traditional DSSE varies from 14.83 to 15.4; whereas, the total error of the proposed DSSE varies from 12.7 to 13.16. To give a better insight, Table 2-2 also includes the percentage of error reduction achieved by the proposed DSSE method over the traditional DSSE method. As one can see, the proposed approach improves the accuracy of DSSE by an average of 13.89%. This suggests that the approach introduced in this chapter has a significant potential to improve the quality of DSSE in distribution systems.
Table 2—2 Statistical Performance of DSSE Error for a Typical Load Profile in the IEEE 13-bus Distribution System

<table>
<thead>
<tr>
<th></th>
<th>Best DSSE (Least total error)</th>
<th>Mean DSSE (Average total error)</th>
<th>Worst DSSE (Most total error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal DSSE</td>
<td>2.6412</td>
<td>2.7714</td>
<td>2.9239</td>
</tr>
<tr>
<td>Traditional DSSE</td>
<td>14.8335</td>
<td>15.0972</td>
<td>15.4202</td>
</tr>
<tr>
<td>Proposed DSSE</td>
<td>12.7026</td>
<td>12.9867</td>
<td>13.1678</td>
</tr>
</tbody>
</table>

Improvement of DSSE accuracy by the Proposed DSSE compared to the Traditional DSSE

<table>
<thead>
<tr>
<th></th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.37%</td>
</tr>
<tr>
<td></td>
<td>13.89%</td>
</tr>
<tr>
<td></td>
<td>14.61%</td>
</tr>
</tbody>
</table>
2.3.1.2 Impact of load uncertainty

The performance of the proposed method is investigated further for a scenario where the load profile has significant unexpected fluctuation. The typical (Scenario I) and unexpected (Scenario II) load profiles are shown in Figure 2-10. The historical data used to update the mean and standard deviation in the following simulation is based on Scenario I. As depicted in Figure 2-10, the load variation is significant and represents a case where the day’s load profile does not follow the pattern of historical data.

![Figure 2-10 Typical and distorted load profiles for 24 hours in pu.](image)

Now, considering Scenario II, the simulation is performed 10000 times, but the same historical data as in the previous section (Scenario I) is used. The results are summarized in
Table 2-3. Although the total error has increased compared to the scenario with a similar/consistent load profile (e.g. the mean of the total error 22.31 compared to 18.95 for the proposed DSSE), the proposed approach still yields good improvement of the accuracy over the traditional DSSE. These results show the robustness of the proposed method. As can be seen in Table 2-3, in scenario II, the average improvement percentage is increased compared to scenario I, because the OD error, which is not considered by the traditional DSSE, is increased as a result of more drastic variation of loads.

<table>
<thead>
<tr>
<th></th>
<th>Best DSSE</th>
<th>Mean DSSE</th>
<th>Worst DSSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal DSSE</td>
<td>2.6312</td>
<td>2.7653</td>
<td>2.9137</td>
</tr>
<tr>
<td>Traditional DSSE</td>
<td>22.0601</td>
<td>22.3125</td>
<td>22.6274</td>
</tr>
<tr>
<td>Proposed DSSE</td>
<td>18.6772</td>
<td>18.9518</td>
<td>19.3143</td>
</tr>
<tr>
<td>Improvement of DSSE accuracy by the Proposed DSSE compared to the Traditional DSSE</td>
<td>15.33%</td>
<td>15.06%</td>
<td>14.64%</td>
</tr>
</tbody>
</table>

2.3.2 IEEE 123-bus distribution system

To evaluate the performance of the proposed method on a larger system, the considered IEEE 123-bus distribution system [100] is implemented in Figure 2-11. In this figure, to evaluate the performance of the proposed method for non-radial systems, the lines in blue are added to create a meshed system. Similarly to the previous system, DSSE is performed 10,000 times for the load profiles given in Figure 2-4 and Figure 2-10. Table 2-4 and Table
2-5 summarize the results of DSSE for this system when using typical load profiles (Scenario I) and highly distorted load profiles (Scenario II), respectively. It can be seen in Tables 2-4 and 2-5 that in both cases, the proposed approach reduces the total error compared to the traditional DSSE by 15.09% and 15.68%, respectively. These results are similar to those obtained for the IEEE 13-bus test system. Thus, the improvement of accuracy achieved by the proposed method is consistent and promising.
<table>
<thead>
<tr>
<th></th>
<th>Best DSSE (Least total error)</th>
<th>Mean DSSE (Average total error)</th>
<th>Worst DSSE (Most total error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal DSSE</td>
<td>12.3209</td>
<td>12.4765</td>
<td>12.9398</td>
</tr>
<tr>
<td>Traditional DSSE</td>
<td>25.5312</td>
<td>25.5590</td>
<td>25.6807</td>
</tr>
<tr>
<td>Proposed DSSE</td>
<td>21.5711</td>
<td>21.7034</td>
<td>22.0211</td>
</tr>
<tr>
<td>Improvement of DSSE accuracy by the Proposed DSSE compared to the Traditional DSSE</td>
<td>15.51%</td>
<td>15.09%</td>
<td>14.25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Best SE (Least total error)</th>
<th>Mean SE (Average total error)</th>
<th>Worst SE (Most total error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal SE</td>
<td>12.3309</td>
<td>12.4680</td>
<td>12.9362</td>
</tr>
<tr>
<td>Traditional SE</td>
<td>39.1400</td>
<td>39.9794</td>
<td>40.6145</td>
</tr>
<tr>
<td>Proposed SE</td>
<td>33.5231</td>
<td>33.7123</td>
<td>33.8581</td>
</tr>
<tr>
<td>Improvement of DSSE accuracy by the Proposed DSSE compared to the Traditional DSSE</td>
<td>14.33%</td>
<td>15.68%</td>
<td>16.64%</td>
</tr>
</tbody>
</table>
CHAPTER 3: INCORPORATION OF DSSE INTO TRANSMISSION LEVEL

3.1 Transmission System State Estimation

As explained in chapter 2, the most established approach for state estimation is the weighted least squares (WLS) method. Unlike distribution systems, transmission systems are assumed to be balanced. Therefore, only the positive sequence system is analyzed. For example, transmission lines are modeled as a two part π model in Figure 3-1 whereas, in the distribution system, they were modeled with the imbalanced 3 phase matrix of the line [100]. Therefore, it is assumed that, in transmission systems, phases are not coupled with each other as only the positive sequence is considered.

![Equation](image)

**Figure 3-1 Equivalent circuit for a transmission line.**

Denoting the total number of buses as \( n \), in an observable distribution system, the rank of the Jacobian matrix is \( 2 \times n - 3 \) since the feeder head bus has three buses with known
angles. However, in transmission distribution systems, the rank of the Jacobian matrix is $2 \times n - 1$ since the transition bus has one bus with known angle.

3.2 Incorporating DSSE Results into TSSE

A schematic diagram of a transmission system interconnected with several distribution systems is displayed in Figure 3-2. From the transmission system point of view, the common bus connecting the two systems is referred to as a PQ (load) bus. From the distribution system viewpoint, this interconnecting point is referred to as the feeder head. In many systems, the power transferred from the transmission to the distribution level through a given PQ bus is measured frequently (e.g., every 4 seconds) using the RTUs installed at substations. This power is used as an injection measurement in TSSE, whose solution will also include the voltage of all the PQ buses (or equivalently, as seen from the distribution side, feeder head estimated voltages).
In this thesis, it is proposed to use the power flow and voltage magnitude at the feeder heads estimated by DSSE as additional power injection and voltage magnitude measurements, respectively, to improve TSSE. As (2-7) indicates, the noise of these additional measurements (estimations) should be represented by their variances. To calculate these variances, (2-11) is rewritten as follows:

$$\Delta x = \left[ G^{-1}(x) \right] \left[ H^T(x) R^{-1} \right] [z - h(x)]$$  \hspace{1cm} (3-1)

where the iteration superscript $k$ has been omitted for clarity. Given that the covariance matrix of the measurements is given by $R$, the covariance matrix $R_x$ of the state vector is
\[
R_x = \left[ G^{-1}(x)H^T(x)R^{-1} \right] R \left[ G^{-1}(x)H^T(x)R^{-1} \right]^T = G^{-1}(x).
\]  

(3-2)

The diagonal elements of $G^{-1}(x)$ are the variances of the estimated bus voltage magnitudes and angles of the distribution system. Hereinafter, the variance matrix of estimated voltage magnitudes of feeder head is referred to as $R_{v_{fh}}$. Similarly to (2-1), the vector of estimated feeder head transferred active and reactive powers, denoted by $z_{PQ_{fh}}$, is related to the state vector through

\[
z_{PQ_{fh}} = y(x).
\]

(3-3)

Next, defining $\begin{bmatrix} \partial y(x) \end{bmatrix}$, (3-3) can be linearized as

\[
\Delta z_{PQ_{fh}} = Y(x) \Delta x.
\]

(3-4)

The covariance matrix of the estimated powers at the feeder head becomes

\[
R_{PQ_{fh}} = Y(x)R_xY^T(x) = Y(x)G^{-1}(x)Y^T(x).
\]

(3-5)

Thereafter, in the proposed TSSE, the diagonal elements of $R_{PQ_{fh}}$ will be used as the variances of the feeder head active and reactive powers calculated based on DSSE.

Let $z_{fh}$ denote a vector consisting of estimated feeder head voltage magnitudes and calculated feeder head powers ($z_{PQ_{fh}}$), and $R_{fh} = \text{diag}[R_{v_{fh}}, R_{PQ_{fh}}]$. According to the proposed approach in
this chapter, the objective function in (2-9) becomes

\[
J = \begin{bmatrix} z_m \\ z_{fh} \end{bmatrix} - \begin{bmatrix} h_m(x) \\ h_{fh}(x) \end{bmatrix}^T \begin{bmatrix} R_m & 0 \\ 0 & R_{fh} \end{bmatrix}^{-1} \begin{bmatrix} z_m \\ z_{fh} \end{bmatrix} - \begin{bmatrix} h_m(x) \\ h_{fh}(x) \end{bmatrix} \\
= \begin{bmatrix} z_{total} - h_{total}(x) \end{bmatrix}^T R_{total}^{-1} \begin{bmatrix} z_{total} - h_{total}(x) \end{bmatrix}
\]

(3-6)

where the subscript “m” refers to the transmission system measurements, the subscript “fh” refers to the feeder head voltage magnitudes and powers which are calculated based on the DSSE results, and the subscript “total” refers to the combination of the transmission system measurements and the additional measurements stemming from DSSE. The solution to the proposed TSSE objective function in (3-6) is similar to the method used in (2-10) and (2-11) for DSSE. The potential improvements from solving the objective function (3-6) are discussed below.

3.3 Areas of Impact in TSSE

3.3.1 Observability

In many practical power systems, the transmission grid may be partially unobservable, particularly on the lower voltages due to insufficient measurements at those levels. This problem is often mitigated by electrical companies by utilizing pseudo measurements at the expense of a decrement in the accuracy and reliability of the results. Generally, pseudo measurements consist of forecasted loads based on historical data. The advantage of using pseudo measurements is that they are inexpensive to obtain without installing new costly measurement devices in the system. In addition to a lack of installed measurement devices,
telecommunication or measuring problems may lead to the unavailability of certain measurements which could also render the system partially unobservable [103], in particular if any of these unavailable measurements are critical [13].

If the measurements of a PQ bus (feeder head) are unavailable due to technical problems or lack of installed measuring devices, the corresponding DSSE could still be executed given that smart meters are installed on all the loads and reclosers are installed along feeders. Once DSSE has been executed and the state of the system is known, a power flow calculation can provide the complex power transferred to the distribution system. This estimated power along with the estimated feeder head voltage magnitude can thus be used as equivalent measurements at the transmission level to restore the observability of the network and achieve a TSSE solution.

3.3.2 Numerical conditioning

Numerical instability (divergence) is another major problem in TSSE [29], which could be a result of lack of redundancy and/or bad initial conditions. A good indicator of numerical stability is the condition number of the TSSE gain matrix –gain matrix is defined in (2-11)– which is calculated as

\[
k_{G} = \|G(x)\|_2 \times \|G^{-1}(x)\|_2
\]  

(3-7)

where \(\|\cdot\|_2\) is the 2-norm of a matrix [103]. Higher condition numbers indicate that the TSSE process will be more sensitive to noise, and therefore, such systems are referred to as ill-
conditioned as they will be more prone to numerical instability. Condition number is function of types and locations of the measurements in a system [29]. As the proposed TSSE provides additional measurements with different types at different locations, the impact of the proposed approach on the condition number should be examined.

### 3.3.3 Bad data identification

Undetected bad data caused by telecommunication issues, measuring malfunctions such as biases or drifts, or malicious attacks may also contribute to the inaccuracy of TSSE. Thus, one of the integral parts of TSSE is to detect bad data, and if possible to exclude them. Therefore, there is a procedure for bad data identification in TSSE program. Some bad data are just eliminated by a simple verification process, e.g. negative voltages and large differences between incoming and leaving currents. Some bad data need to be treated more carefully through a rather complicated process. If there is sufficient redundancy in a system, such bad data are expected to be filtered. For this purpose, both pseudo measurements and DSSE results could provide the required redundancy. However, pseudo measurements tend to be very inaccurate to the extent that they could themselves become bad data and complicate the bad data identification procedure. Since measurements based on DSSE results as presented in chapter 2 are more reliable, they have the potential to improve the ability of existing algorithms to identify and remove the bad data.

It is important to mention that bad data detection methods are only effective against non-critical measurements [13]. Therefore, if measurements of a PQ bus are critical and include bad data, they are not detectable regardless of the magnitude of the error [13]. Since the
DSSE results could render those bad measurements non-critical, the proposed approach may also enable identifying and eliminating them.

There have been many studies on bad data identification in power systems, e.g., [13], [19], [20], [23], [24]. In this section, the normalized residual method is employed [13], [23]. This method is based on the properties of measurement residuals. Consider the linearized measurement equation (2-1) in the following form [25]

\[ \Delta z = H(x)\Delta x + v \]  \hspace{1cm} (3-8)

The WLS estimation will give the state vector as

\[ \Delta \hat{x} = \left[ H^T(x)R^{-1}H(x) \right]^{-1} H^T(x)R^{-1}\Delta z = G^{-1}(x)H^T(x)R^{-1}\Delta z \]  \hspace{1cm} (3-9)

where \( \Delta \hat{x} \) is the estimate of the state vector. Then, the estimated vector of measurements is

\[ \Delta \hat{z} = H(x)\Delta \hat{x} = K\Delta z \]  \hspace{1cm} (3-10)

where \( \Delta \hat{z} \) is the calculated measurements based on the estimate of the state vector and

\[ K = H(x)G^{-1}(x)H^T(x)R^{-1} \]. The measurement residuals can be expressed as follows:

\[ r = \Delta z - \Delta \hat{z} = [I - K]\Delta z = S\Delta z \]  \hspace{1cm} (3-11)
where \( I \) is the identity matrix, and \( S = I - K \). To obtain the vector of normalized residuals \( r^N \), the absolute value of each residual is divided by the square root of the corresponding diagonal element of \( S \) as in

\[
    r_i^N = \frac{|r_i|}{\sqrt{S_{ii}}}
\]

(3-12)

where the subscript \( i \) refers to the \( i \)th measurement. When a normalized residual is greater than a specific threshold, its corresponding measurement is assumed to be incorrect. Some utilities eliminate the measurement with the biggest normalized residual (above a certain threshold) and perform the TSSE again recursively until all normalized residuals are below the assumed threshold.

### 3.3.4 Accuracy

Due to the use of finite precision arithmetic in computers, the ill-conditioned system may converge to a point away from the optimal solution. Moreover, the normalized error of TSSE is bounded by the condition number as [103], [104]

\[
    \frac{\|\hat{x} - x\|_2}{\|s\|_2} \leq \frac{\|v\|_2}{\|z\|_2} \left( \sqrt{k_G \left( \frac{\|z\|_2}{\|H(x) - z\|_2} + 1 \right)} + k_G \frac{\|H(x) - z\|_2}{\|H(x)\|_2 \|x\|_2} \right).
\]

(3-13)

where \( v \) is the noise vector introduced in (2-1), \( H(x) \) is the Jacobian matrix defined in (2-10), \( z \) is the measurement vector, \( x \) is the actual/ideal state of the system without any errors, \( \hat{x} \) is the estimate of the state, and \( k_G \) is the condition number of the gain matrix. It
has been shown that when $k_c$ rises, the normalized error also rises [103]. Therefore, by improving the numerical conditioning of TSSE using DSSE results, more accurate estimation at the transmission level can be expected.

Generally, the accuracy of measuring devices improves as technology progresses. For example, the recently installed measuring devices, e.g. smart meters, are generally more accurate than RTUs which were installed decades ago. Consequently, as utilities keep on improving their distribution system measuring infrastructure by installing more smart meters while also refining their modeling database, it is expected that DSSE will be able to provide accurate estimates of the state of the distribution system (including voltage magnitudes of feeder heads) and the power transferred through the PQ buses. These estimates will then replace the less reliable pseudo measurements, thus increasing the TSSE accuracy.

### 3.4 Case Studies

To demonstrate the proposed approach, this section presents studies in which the results of DSSE are incorporated into TSSE using two transmission benchmark systems. The two transmission systems are the IEEE 14-bus system and the IEEE 57-bus system [105]. In each case study, the transmission system is connected to a distribution system at every PQ bus. The considered distribution system is the IEEE 13-bus system [100]. Given that the distribution systems are typically unbalanced, three-phase component models are considered in each distribution system. Network parameters and load base values are
modified to ensure a feasible solution with voltages in acceptable range exists for different load levels.

In the first case study, with the IEEE 14-bus transmission system, the overall combined system contains 330 single-phase buses. Figure 3-3 demonstrates the topology of this system. In Figure 3-3(a), the IEEE 14-bus transmission system is shown. In Figure 3-3(b), the connection of one typical PQ bus of the transmission system (bus 9) to a distribution system is shown. The transition buses in Figure 3-3(b) are considered the border line between modeling transmission and distribution systems. If a PQ bus in Figure 3-3(a) has measurement devices, it means the active and reactive powers are measured on the secondary side of the corresponding transformers connected to it. In the second case study, with the IEEE 57-bus transmission system, the overall system has 1771 single-phase buses.

To create more realistic case studies, it is assumed that the network parameters have some degree of uncertainty. The uncertainty of network parameters is considered using a uniform distributed random variable [102]. Bounds of uncertainty of 2% [106] for transmission systems and 5% for distribution systems are considered.

The measurements noise is modeled as a normally distributed random variable with a specific standard deviation. The standard deviation for each measurement is assigned based on the precision of the measuring device as
Figure 3—3 Configuration of the first case study: (a) Modified IEEE 14-bus transmission system with allocated loads, generators, and measurement devices; (b) The feeder head of the IEEE 13-bus distribution system is connected to bus 9 of the IEEE 14-bus transmission system

\[ \sigma = \frac{pr \times |z|}{3} \]  

(3)

where \( pr \) is the device precision, and \( |z| \) is the absolute of the true value of the measured electric variable. Typical power and voltage magnitude measurement precisions of 3% and 1.5%, respectively, are assumed for transmission systems [7].
A daily load profile measured for a building at The University of British Columbia is given in Figure 3-4 (a) [91]. During a 2-hour interval, the given building load follows a pattern similar to Figure 3-4 (b), which is extracted from Figure 3-4 (a) for the interval from 2 pm to 4 pm.

To demonstrate the proposed approach, herein it is assumed that TSSE is executed every 60 seconds. The smart meters are assumed to measure every 15 minutes [12]. Moreover, to emulate a realistic scenario as explained in [12], [107], the smart meters are not synchronized, implying that at any given time during the simulation some measurements have just been updated while the others are out-of-date by up to 15 minutes as explained in chapter 2.

Figure 3–4 Typical 24 hour load profile normalized with respect to the peak value based on historical data recorded at The University of British Columbia [91]; (b) Recorded 2-hour fragment of load profile from 2pm to 4pm.
The system is simulated for a 2-hour interval using one-minute time steps. At each step of the simulation, first the loads are updated according to Figure 3-4 (b). Afterwards, a 3-phase unbalanced optimal power flow [108] is performed in order to obtain the generators’ dispatch and get a reference solution in accordance to the status of all loads. After that, Gaussian noise is added to the calculated values of active and reactive powers and voltage magnitudes in order to emulate the measurements errors. Since the smart meters are not synchronized, only some of their measurements are up-to-date at any given step of the simulation, while the rest are out-of-date as explained in [12].

Two methods are implemented for comparison. The procedure of calculating the TSSE results at each time step using the traditional and proposed methods is depicted in Figure 3-5. In the first method, herein referred to as the Traditional TSSE approach, TSSE is executed by itself without any extra information from the distribution system. In the second method, referred to as the Proposed TSSE approach, both DSSE and TSSE are executed according to the methodology presented in Section 2.2.

3.4.1 Restoring observability

Before performing TSSE in a system, the observability of the system should be checked since measuring and communication problems may render parts of the system unobservable on a regular basis [103]. In Figure 3-3 (a), the measurement devices in the transmission system are presented. Based on this allocation of measurement devices, the observability analysis [13] of the system depicted in Figure 3-3 (a) shows that the measurements of two device are critical in this system; the device at bus 9 and the device
between bus 4 and bus 7. The absence of any other measurement will not stop the execution of TSSE; however, if any of the two aforementioned devices fail, the gain matrix in (2-11) will become singular and the TSSE execution will stop.

Figure 3–5 Procedure of calculating the results of Proposed and Traditional TSSE at each step of the simulation.

To demonstrate the improvement of observability, in the first scenario, it is assumed that the measurement device of bus 9 and/or its communication system has failed. Since this is
a critical measurement, the transmission system is not fully observable and TSSE cannot be successfully executed. However, since smart meters are installed at each load in Fig 3-3 (b) and reclosers have voltage magnitude and power measurement devices, the DSSE can be executed successfully at any time. Based on DSSE results, the voltage magnitude and injected active and reactive powers at the distribution system feeder head and their variances are calculated according to section 3.2. As the feeder head is connected to bus 9 as shown in Fig 3-3 (b), following the proposed approach in (3-10), the calculated powers at the feeder head will restore the observability of the transmission system.

In the second scenario, it is assume that the measurement device between bus 4 and bus 7 and/or its communications system has failed. Unlike scenario one, in this case, DSSE based calculation of none of the distributions systems will be a replacement for the missing measurements. However, a great property of the proposed TSSE is that, even in this scenario, observability can be restored. As Figure 3-6 shows, the injected power to the region encompassing bus 4 to bus 7 transmission line is available by the red measurement devices. Given the DSSE results in PQ buses 4, 5, 8, 9, 13, and 14, and assigning the generator in bus 2 as the slack bus, the injection power at all buses is either measured or calculable in the depicted region. In a system where all of the injection powers are available, conventional load flow can be executed and bus voltages can be calculated. This means that the system is observable and the proposed TSSE can obtain the system state and maintain the observability of the system despite that a critical measurement has failed.
Figure 3–6 IEEE 14-bus transmission system with a faulty critical measurement device. The subsystem which has a faulty measurement device is depicted. Red measurement devices show the injection power from the rest of the system to the subsystem.

3.4.2 Improving numerical conditioning

In the next set of studies it is assumed that all measurement devices are functioning properly. To demonstrate one of the additional benefits of incorporating DSSE results in TSSE, the condition numbers of the TSSE gain matrix for the two benchmark systems have been calculated during the 2-hour study interval. Since the condition numbers also change with the measurements, the calculated maximum, minimum, and mean condition numbers obtained during the 2-hour interval are summarized in Table 3-1. As Table 3-1 shows,
employing the DSSE results as additional measurements (by 2 orders of magnitude) improves the condition numbers for both benchmark systems. In practice, one can further improve the numerical characteristics using numerical conditioning methods such as orthogonal transformation [13]. Here, to show the impact of the proposed method on numerical characteristics exclusively, other numerical methods are not added to the TSSE formulation.

**Table 3—1 Statistical Results of Calculating the Condition Number of TSSE for 100 Runs**

<table>
<thead>
<tr>
<th>Benchmark system based on the IEEE 14-bus transmission system</th>
<th>Method</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Traditional TSSE</td>
<td>3.6423e+007</td>
<td>2.8646e+011</td>
<td>8.4949e+11</td>
</tr>
<tr>
<td></td>
<td>Proposed TSSE</td>
<td>8.0149e+006</td>
<td>1.5899e+09</td>
<td>2.3343e+09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benchmark system based on the IEEE 57-bus transmission system</th>
<th>Method</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Traditional TSSE</td>
<td>2.7879e+09</td>
<td>5.7949e+011</td>
<td>9.1302e+11</td>
</tr>
<tr>
<td></td>
<td>Proposed TSSE</td>
<td>3.9346e+08</td>
<td>6.2415e+09</td>
<td>6.5149e+09</td>
</tr>
</tbody>
</table>

The number of iterations to achieve a converged solution is another indicator of the numerical stability of the TSSE problem. Figure 3-7 shows the recorded number of iterations it took to converge in each case study. In the first study with a smaller IEEE 14-bus benchmark system [see Figure 3-7 (a)], the proposed approach reduces the number of iterations from a range of 4 to 8 to a range of 4 to 5 iterations. Similarly, Figure 3-7(b), shows the recorded number of iterations for the IEEE 57-bus transmission system. Here,
the difference between Traditional and Proposed TSSE approaches is even more pronounced. As it can be seen in Figure 3-7(b), the number of iterations using the Traditional TSSE often between 10 and 20, and there are two points where the method fails to converge after 150 iterations. However, the Proposed TSSE approach has always converged and remained within a reasonable range of 4 to 8 iterations.

(a)
Figure 3–7 Number of iterations taken to achieve a converged TSSE solution for: (a) the IEEE 14-bus transmission system; and (b) the IEEE-57 bus transmission system.

The total and average numbers of iterations for these two cases (over the two hours) are summarized in Table 3-2. As it can be seen from Table 3-2, the Traditional TSSE method required 3,678 and 8,864 iterations for the small and large benchmark systems, respectively, whereas the Proposed TSSE method took 3,363 and 3,429 iterations, respectively, for the same two systems. On the other hand, the size of the optimization problem in (10) is increased by about 60% because of the additional measurements at the PQ buses, which imposes extra computations. Therefore, it is possible that the computational burden is increased in total. In practice, utilities such as BC Hydro often use pseudo measurements for all loads. The proposed method, once implemented, would only replace pseudo measurements with feeder head estimations. Thus, practically, the computational burden would not increase as a result of the proposed method.
<table>
<thead>
<tr>
<th>Benchmark system based on the IEEE 14-bus transmission network</th>
<th>Method</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional TSSE</td>
<td>613</td>
<td>5.11</td>
<td></td>
</tr>
<tr>
<td>Proposed TSSE</td>
<td>560</td>
<td>4.67</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benchmark system based on the IEEE 57-bus transmission network</th>
<th>Method</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional TSSE</td>
<td>1477</td>
<td>12.31</td>
<td></td>
</tr>
<tr>
<td>Proposed TSSE</td>
<td>619</td>
<td>5.16</td>
<td></td>
</tr>
</tbody>
</table>

### 3.4.3 Impact on accuracy of TSSE

Taking into consideration more measurements (from DSSE) with higher accuracy as well as an improved condition number should further improve the accuracy of TSSE. To examine the impact of the Proposed TSSE approach on TSSE accuracy, the total error of the estimated voltages over all buses is calculated at every time step as

\[
\text{error} = \sqrt{\sum_{i=1}^{\text{no\_bus}} \left( v_i^{\text{est}} - v_i^{\text{act}} \right)^2}
\]  

where \(\text{no\_bus}\) is the number of buses, and \(v_i^{\text{est}}\) and \(v_i^{\text{act}}\) are the estimated and exact (errorless) voltages, respectively, at the \(i\)th bus. Each voltage variable in (3-15) is a complex number in per unit. The accumulative error in (3-15) over the 2 hour simulation interval is referred to as the total error of the estimation.
Figures 3-8(a) and 3-8(b) demonstrate the calculated error achieved by the Traditional and Proposed TSSE approaches compared to the traditional TSSE for the IEEE 14-bus and 57-bus systems, respectively. In Figure 3-8(a), it is clear that the error has significantly decreased both in terms of its value at every step as well as its average over the 2-hour interval. Similar observations are made in Figure 3-8(b) for the larger system. These improvements are due to a number of factors such as using DSSE results and better numerical conditioning of the problem. However, Figure 3-8(b) also shows several spikes in the error of the Proposed TSSE method, which is the result of using out-of-date measurements used in DSSE.
The accuracy of DSSE depends on many factors including the variation of loads, the accuracy of the smart meters, as well as the rate at which the smart meters measurements are updated. To study the significance of smart meters measurement updating rate, another simulation is executed on the IEEE 57-bus test system assuming that their rate is increased from 15 to 30 minutes. The results are presented in Figure 3-9. As this figure shows, a reduction of the sampling rate noticeably increases the error of the estimation at certain times. The reason is that some loads may vary significantly between two subsequent measurement updates of the smart meters, thereby yielding larger errors between these two points. The probability and magnitude of this variation increases with the duration of the sampling interval. However, it is important to point out that the system
still converges for all steps, and the numerical conditioning is still improved. Figure 3-9 shows that it is beneficial for utilities to invest on increasing the resolution of smart meter measurement updates in their system.

![Error of TSSE](image)

**Figure 3-9** Total error of TSSE evaluated over all buses at each time step and its average for the 2-hour study interval when the smart meters measurements are updated every 30 minutes for the IEEE 57-bus transmission system.

### 3.4.4 Facilitating bad data identification and accuracy improvement

Improving the identification of anomalies in TSSE is another important aspect of the Proposed TSSE approach. The effectiveness of bad data identification methods depends on many factors, including the level of redundancy and the accuracy of the measurements. In order to demonstrate this contribution, in this section, TSSE of the IEEE 14-bus benchmark system at 2 pm is considered. To emulate a bad datum, the active power injection...
measurement at bus 10, which is a non-critical measurement, is distorted to contain 10% error. This error could be a result of false-data injections or technical problems. The calculated initial normalized residuals (i.e., before bad data detection) of the active and reactive power injections at bus 10 using the Traditional and Proposed TSSE approaches are summarized in Table 3-3.

<table>
<thead>
<tr>
<th>Bus</th>
<th>Type of Measurement</th>
<th>Normalized Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>active power</td>
<td>2.2905</td>
</tr>
<tr>
<td>10</td>
<td>reactive power</td>
<td>0.7462</td>
</tr>
</tbody>
</table>

Based on the normal distribution characteristics, a threshold of 2.57 is assumed to determine the presence of bad data with 95% confidence [7]. As it can be seen in Table 3-3, with the Proposed TSSE approach, the residual of the active power measurement of bus 10 is greater than 2.57. However, with Traditional TSSE, the normalized residual is below the assigned threshold. The reason is that the proposed method provides more redundancy to identify bad data. At the same time, the measurement of reactive power (which is not corrupted) is correctly classified as regular data. The benefit of removing bad data is shown in Table 3-4, wherein the total error [see (16)] of Proposed TSSE is shown to decrease from 0.0393 to 0.0346 when the erroneous power measurement is removed. Table IV also demonstrates the improvements in accuracy obtained by using Proposed TSSE before removing any bad data in comparison to Traditional TSSE. The proposed approach also
enables detection of bad data in buses with critical active and reactive power injection measurements. In Traditional TSSE, the residuals of critical measurements are always zero, whereas the proposed approach renders those measurements non-critical by providing redundant measurements resulting in non-zero residuals.

Table 3–4 Total Error of TSSE for the IEEE 14-bus Transmission System with Bad data at Bus 10 and Its Elimination

<table>
<thead>
<tr>
<th>Traditional TSSE</th>
<th>Proposed TSSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0.0588</strong></td>
<td><strong>0.0346</strong></td>
</tr>
<tr>
<td><strong>0.0393</strong></td>
<td><strong>0.0346</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No Bad Data Identification</th>
<th>Applying Bad Data Identification and Elimination</th>
</tr>
</thead>
</table>
CHAPTER 4: BAD DATA IDENTIFICATION AND ELECTRICITY THEFT

4.1 Modeling Theft as Bad Data

Sometimes in power systems, typically at the distribution level, there are a number of users who are stealing electricity by bypassing their measurement devices. A simplified diagram depicting connection of smart meters at the distribution feeder for modeling electricity theft and bad data is shown in Figure 4-1. Specifically, in figure 4-1(a), load 2 is partially bypassing its smart meter and stealing electricity. Therefore, the quantities which are reported by the smart meter will become erroneous. From the point of view of received information, this case is analogous to a situation where there is no manipulation with the metering device, but the meter is malfunctioning, as depicted in Figure 4-1(b). In this case, there is no bypassing (stealing) at load 2, but its smart meter is malfunctioning and it may be reporting the same quantity as in Figure 4-1(a). Sometimes, despite no bypassing, the user tampers the meter to report less consumption and steal electricity. These cases of tampering the meters are also analogous to cases of malfunctioning devices. In both cases (theft and/or meter malfunction), the delivered power from the system to the load and the voltage at the load are the same since this measurement does not impact the operation of the physical system.
Figure 4—1 A simplified diagram depicting connection of smart meters for modeling electricity theft as bad data: (a) load 2 is partially bypassing its smart meter to commit electricity theft; and (b) smart meter of load 2 is malfunctioning.

The measurements of malfunctioning devices are referred to as bad data. Therefore, an algorithm to identify bad data would also help to locate the loads which are potentially stealing electricity. In order to distinguish between the bad data due to device malfunctioning or due to the theft, three general criteria could be considered. The first criterion is recognition of obvious bad data. If bad data are obvious, such as negative measured voltage magnitude values or the measurements that are clearly out of range, it is more likely that the device is actually malfunctioning (as the thieves would generally manipulate their meters in a non-obvious way [16]). The next criterion is when the reported measured power is more than its actual value. These cases are certainly not instances of theft. However, determining the user in these cases is important and helpful to utilities since the user is being billed more than they are consuming. Sometimes, a device is
malfuctioning due to a temporary problem such as telemetry failures. Usually, in these cases, the device only malfunctions for a short interval. However, a tampered or bypassed device would consistently transfer bad data for an extended period. In this chapter, in the first stage, the problem of bad data identification to locate the two types of theft points which result from meter tempering and smart meter bypassing is formulated as an optimization problem.

4.2 Problem Formulation of Bad Data Identification

The normalized residual method for bad data identification has been presented in Section 3.3.3. This method is most efficient to identify single bad data in a set of measurements. However, for multiple bad data, it could classify some good measurements as bad data or miss bad data [25]. In this section, a method to identify multiple bad data is presented.

Identification of multiple bad data is treated as an optimization problem based on the Decision Theory approach [70]. For a set of $m$ measurements, the decision as to whether the $i$th measurement is good or bad is represented using a binary variable, $d_i$. Based on $d_i$, a measurement considered as good or bad data as follow;

$$d_i = 0, \quad \text{if the } i\text{th measurement is bad data}$$

$$d_i = 1, \quad \text{if the } i\text{th measurement is good data}.$$

A decision vector is defined as a combination of decisions (binary variables) on a set of $m$ measurements, as $d = [d_1, d_2, d_3, ..., d_m]$. Thus, if the system is fully observable, $2^m$ possible
decision vectors exist. Each decision vector referred to as a plausible for the observed data if, after the removal of the measurements that have been labeled as bad data, no more bad data is detected.

In order to detect whether or not bad data exists in a set of measurements, a hypothesis test should be used [69]. The DSSE objective function as defined in Chapter 2 is

$$J(x) = \sum_{i=1}^{m} \left( \frac{z_i - \mu_i}{\sigma_i} \right)^2.$$ \hspace{1cm} (4-1)

As (4-1) shows, $J(x)$ is the summation of the square of $m$ normally distributed random valuables. Let $n$ denote the number of state variables (which is the size of $x$ vector). Given that $m$ measurements are used in (4-1), according to the hypothesis test, $J(x)$ is a random variable which follows the chi-square distribution with $m-n$ degrees of freedom [13], denoted as $\chi^2_{m-n}$. For example, Figure 4 – 2 shows the probability density function (pdf) of the chi-square distribution with 8 degrees of freedom i.e., $m-n = 8$. The null hypothesis is defined as the assumption that there are no bad data in the set of measurements. Therefore, as the only source of error to measurements would be the Gaussian noise, the error of the $i$th measurement in (4-1) follows a normal distribution with mean equal to zero and variance equal to $\sigma_i$. The considered test verifies whether or not the null hypothesis is a correct assumption. To do so, the value of $J(x)$ is compared against a pre-calculated threshold, which is depicted in Figure 4 – 2 as $C$. This threshold is defined in a
manner that if \( J(x) \) is greater than it, the null hypothesis is rejected, and if \( J(x) \) is smaller than it, the null hypothesis is accepted.

![Probability density function of the chi-square distribution with 8 degrees of freedom.](image)

**Figure 4–2 Probability density function of the chi-square distribution with 8 degrees of freedom.**

The value of threshold \( C \) corresponds to the confidence level in hypothesis test, denoted as \( \alpha \). For example, if hypothesis test is set to be correct for 95% of the time (\( \alpha =0.95 \)), the threshold \( C \) would be calculated from the following;

\[
\int_0^C \varphi(\tau) d\tau = \alpha = 0.95 \quad \text{with} \quad \varphi(\tau) = \frac{(m-n)^{-1/2} e^{-\tau/2}}{2^{(m-n)/2} \Gamma((m-n)/2)}
\]  

(4-2)

where \( \varphi(\tau) \) is the \( pdf(\chi_{m-n}^2) \); and \( \Gamma \) is the Gamma function [25]. According to hypothesis test, the null hypothesis is accepted {meaning that the measurements \([z_i, z_2, \ldots, z_m]\) in (4-1) are considered as good data} only if \( J(x) \) is less than the threshold \( C \).
The decision vector $d$ contains the information on which measurements are considered to be bad. By removing bad measurements from (4-1), there will be $m_{\text{remained}}$ remaining measurements, denoted as $z(d)$. Then, DSSE is executed using the remaining measurements $z(d)$, the result of which is the corresponding state variables $x(d)$ and the value of objective function $J(x(d))$. The remaining $m_{\text{remained}}$ (good) measurements are used in (4-2), based on $m_{\text{remained}} - n$ degrees of freedom, to calculate the corresponding threshold $C(d)$. Then, the value of $J(x(d))$ is re-compared with $C(d)$ to complete the Hypothesis test process as

$$\begin{align*}
J(x(d)) &> C(d), & z(d) \text{ contains some measurements that are bad data} \\
J(x(d)) &< C(d), & \text{All measurements in } z(d) \text{ are assumed as good data}
\end{align*}$$

If $z(d)$ passes the hypothesis test, the decision vector $d$ is considered plausible. Generally, more than one plausible decision vector. Therefore, a plausible decision vector with maximum likelihood should be found.

Let $p_i$ denote the probability of measurement $i$ to be good, and $q_i$ denote the probability of it to be bad data. A decision vector separates the measurements into good and bad data. If $G$ is the set of good measurements, and $B$ is the set of bad measurements, then the probability of a given decision vector $d$ would be
\[ P(d) = \prod_{i \in G} p_i \prod_{i \in B} q_i. \]  \hfill (4-3)

Generally, it is assumed that all measurement devices have the same reliability. Therefore, all measurements are assumed have the same \( p_i \) and \( q_i \). Given that \( p_i \) is close to one [25], we have

\[
\log(P(d)) = \sum_{i \in G} \log(p_i) + \sum_{i \in B} \log(q_i) \approx \sum_{i \in B} \log(q_i) = \sum_{i=1}^{m} (1 - d_i) \log(q_i) = \log(q_i) \sum_{i=1}^{m} (1 - d_i). \]  \hfill (4-4)

In (4-4), the log function is used to simplify the equation. The decision vector with the most likelihood maximizes \( \log(P(d)) \). Therefore, the solution which maximizes \( \log(P(d)) \) would also minimize the objective function \( \Psi(d) \) as

\[
\text{Minimize } \Psi(d) = -\log(P(d)) = -\log(q_i) \sum_{i=1}^{m} (1 - d_i) \equiv \text{Minimize } \Psi(d) = \sum_{i=1}^{m} (1 - d_i). \]  \hfill (4-5)

This means that the most likely decision vector is the one which results in identifying the minimum number of bad measurements.

However, a decision vector is considered unacceptable if the resulting system is not observable. Also, it must pass the hypothesis test. Therefore, the identification of multiple bad data problem is formulated as
Minimize $\Psi(d) = \sum_{i=1}^{m} (1 - d_i)$

subject to:  

System is observable  

$J(x(d)) < C(d)$

### 4.3 Solution Method

Various optimization techniques have been proposed to solve optimization problems including linear programming (LP), and its hybrid versions, Newton-like methods, nonlinear programming (NLP), quadratic programming (QP), interior point method (IPM), sequential unconstrained minimization [109]-[111], etc. Most of these classical optimization methods are limited to objective functions and constraints that are convex, differentiable, and continuous, and may depend on specific functions and/or constraints [112]. Due to the nature of these methods, they might converge to local solutions and fail to achieve the global optimum for non-convex problems with complicated constraints [113]. For example, it is very difficult to incorporate the observability constraint in (4-6) into classic optimization methods. A new category of optimization tools including evolutionary algorithms (EAs) [114]-[117], Simulated Anealing (SA) [118], Artificial Neural Network (ANN) [119]-[122], Tabu Search Algorithm (TSA) [123], dual-type methods [124], [125], mean field theory [126], ordinal optimization theory [108], etc. have been proposed. In this chapter, a new solution for multiple bad data identification problem based on shuffled frog leaping algorithm (SFLA) optimization is proposed. The SFLA is a meta-heuristic optimization method based on observing and modeling behavior of frogs. The SFLA combines the benefits of the genetic-based mimetic algorithms (MAs) and the social
behavior-based particle swarm optimization (PSO) algorithm [127], [128]. Since this algorithm has not been applied in multiple bad data identification problem in the past, to demonstrate the capabilities, the results of it are also compared to GA and PSO methods which have been previously applied to solve this problem.

### 4.3.1 Shuffled frog leaping algorithm

In SFLA, there is a population of possible solutions defined by a set of frogs that is divided into subgroups called “memplexes,” each performing a local search. At first, an initial population of $F$ frogs is created randomly within the feasible search space. For an $m$ variable optimization problem, the $i$ th frog is represented as $X_i = [x_1, x_2, \ldots, x_m]$. The fitness of each frog is its value of the objective function. After calculating the fitness of the frogs, they are sorted in a descending order according to their fitness. Then, the whole population of $F$ frogs is divided into $S$ memplexes, each containing $\frac{F}{S}$ number frogs. In this procedure, the first frog moves to the first memplex, the second frog moves to the second memplex, and so on all the way until the $S^{th}$ frog moves to the $S^{th}$ memplex. Then, the $(S + 1)^{th}$ frog goes to the first memplex, $(S + 2)^{th}$ frog goes to the second memplex, and so on until all of the frogs are distributed in memplexes.

Within each memplex, the frogs with the best and worst fitness are labeled as $X_p$ and $X_w$, respectively. Also, the frog with the best global fitness is denoted as $X_g$. Then, in each memplex, a procedure is applied to improve the frog with the worst fitness as follows:
\[ L_i = \text{Rand}(0,2) \times (X_b - X_w). \] \hfill (4-7)

\[ X_{w_{\text{NEW}}} = X_w + L_i. \] \hfill (4-8)

where \( \text{Rand}(0,2) \) is a vector of random numbers between 0 and 2, \( L_i \) is the leap size, and \( X_{w_{\text{NEW}}} \) is the new position of the worst frog after leaping. If this procedure does not result a better solution, the calculations in (4-7) and (4-8) are repeated with replacement of \( X_b \) by \( X_g \). These calculations constitute to the next attempt. If no improvement is achieved in this case either, then a new frog is randomly generated within the feasible search space to replace the worst frog. These calculations are repeated for a specific number of iterations for each memeplex. Afterwards, to ensure global exploration, the memeplexes are shuffled, and the frogs are sorted and distributed among memeplexes again [129]. The local search within memeplexes and the shuffling continue until convergence criteria are satisfied.

### 4.3.2 Satisfying the constraints

A common approach to satisfy constraints in optimization problems is to modify the objective function using penalty factors [115], [130]. Since the goal of optimization is to minimize the objective function, violation of constraints in (4-6) is formulated as additional terms to the objective function using penalty factors. Therefore, the objective function in (4-6) is modified as
Minimize $\Psi(d) = -\sum_{i=1}^{m} (t - d_i) + \rho_{\text{Plause}} + \rho_{\text{Observe}}$. \hspace{1cm} (4-9)

where $\rho_{\text{Plause}}$ and $\rho_{\text{Observe}}$ are the plausibility and observability penalty terms defined as:

$$\rho_{\text{Plause}} = \begin{cases} 0 & \text{if } J(x(d)) < C(d) \\ k_1 \times (J(x(d)) - C(d)) & \text{if } J(x(d)) > C(d) \end{cases}$$

$$\rho_{\text{Observe}} = \begin{cases} 0 & \text{if the system is observable} \\ k_2 & \text{if the system is unobservable} \end{cases}$$ \hspace{1cm} (4-10)

where $k_1$ and $k_2$ are the penalty factors.

### 4.3.3 Implementation of SFLA as the proposed solution

To use SFLA for multiple bad data identification, each frog corresponds to a decision vector. In a system with $n$ state variables and $m$ measurements, in order to maintain observability, at most $m-n$ measurements could be eliminated. Therefore, the defined frog in the optimization process is a $m-n$-dimensional vector. Arrays whose value, $l$, is a number from 1 to $m$ indicates elimination of the $l$th measurement. Since $m-n$ is the maximum number of measurements to be eliminated, some arrays may not indicate elimination of any measurements. The arrays whose values are greater than $m$ or less than zero are interpreted as not eliminating any measurements. Therefore, the boundaries on an array are set to $[-m, 2m]$ interval to enable SFLA to explore all possibilities including not eliminating any measurements. During the optimization process, whenever a new frog is generated, each array of it is checked to see whether it is within the $[-m, 2m]$ interval.
boundaries. If an array violates a boundary, its value is adjusted to be equal to the violated boundary.

Since SFLA was originally designed for continuous search spaces, some modifications are required for a discrete search space since the optimization output must be a vector (frog) with integer values. Here, in the considered implementation, the search space is considered continuous and the arrays of frogs are not defined as integer values. However, to evaluate the fitness of a frog, first the values of arrays are temporarily rounded to their nearest integer. Then the objective function corresponding to the frog is evaluated. Afterwards, the arrays of the frog will be reverted to the original values before rounding. Therefore, next population of frogs is generated from the original values of arrays (before rounding), and the rounded values are only used to evaluate the fitness of frogs based on (4-9). Figure 4-4 summarizes the optimization process.
4.3.4 Case study

To demonstrate the proposed methodology, the IEEE 123-bus distribution system [100] as depicted in Figure 4-4 is considered. Among the modifications, the lines in blue are added
to create some loops in the system with single phase loads. The buses are numbered in a manner that the difference between bus numbers is proportional to the distance between two buses in general. For the purpose of this study, a total of eight points of theft have been assumed, which are highlighted as read dots in Figure 4-5. Voltage magnitude measurements of smart meters are not used in this stage. The amount of stolen energy varies from one theft-load point to another. Without loss of generality, the percent of stolen power, which is not measured by the corresponding smart meter at each theft point, is summarized in Table 4-1. For example, the total consumption of load 7 in Table 4-1 will be the measured power multiplied by $\frac{1}{1-0.7}$.

![Figure 4-4 Considered IEEE 123 distribution systems depicting eight loads with electricity theft.](image)
Table 4–1 Percentage of the Unmeasured Consumption at Each Theft Point

<table>
<thead>
<tr>
<th>Theft load point</th>
<th>7</th>
<th>16</th>
<th>41</th>
<th>62</th>
<th>69</th>
<th>70</th>
<th>71</th>
<th>84</th>
<th>85</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of consumption that is not reported in measurement</td>
<td>70</td>
<td>20</td>
<td>95</td>
<td>35</td>
<td>55</td>
<td>40</td>
<td>40</td>
<td>50</td>
<td>60</td>
</tr>
</tbody>
</table>

In addition to SLFA, genetic algorithm (GA) [131] and PSO [132] have also been implemented to solve the multiple bad data identification problem. The number of frogs in SFLA, particles in PSO, and chromosomes in GA is set to 60. In SLFA, the number of memeplexes is set to 6. The confidence level parameter in (4-2) is set to 95%. The penalty factors $k_1$ and $k_2$ in (4-10) are set to 1 and 10, respectively.

The results of the simulation are shown in Table 4-2. As the Table 4-2 shows, both the GA and PSO methods found only five theft points (missed theft four points!) while satisfying the observability constraint. As Table 4-2 shows, the identified theft points in these two methods are different while their DSSE objective function, $J(x)$, has the same value. The reason is that eliminating the measurements of one of the three theft points, 69, 70, or 71 will render the other two as critical measurements. As a result, the residual of the two remaining measurements in DSSE would be zero. Therefore, whether theft point 70 is eliminated in GA or theft point 69 is eliminated in PSO, the WLS objective function would have the same value. This means that all of the frogs which satisfy the constraints in (4-6) at the end of the optimization should be determined, and the measurements they identify should be considered as suspicious theft point. Among the three applied methods, the SFLA
happens to be the most successful one by finding six theft points and missing only three points, while satisfying the observability constraint. The reason that these three theft points are not identified is that they become critical measurement after eliminating the previously identified theft points. Therefore, their residual becomes zero and they would not have any impact on $J(x(d))$, which makes them unidentifiable. Moreover, if the theft amount is very small and temporary, it is possible not to be identified even when there is redundancy of measurements.

<table>
<thead>
<tr>
<th>Solution Method</th>
<th>Identified Theft Points</th>
<th>Unidentified Theft Points</th>
<th>$J(x(d))$</th>
<th>Observability Constraint Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>7, 41, 62, 70, 85</td>
<td>16, 69, 71, 84</td>
<td>46.725</td>
<td>Satisfied</td>
</tr>
<tr>
<td>PSO</td>
<td>7, 41, 62, 69, 85</td>
<td>16, 70, 71, 84</td>
<td>46.725</td>
<td>Satisfied</td>
</tr>
<tr>
<td>SFLA</td>
<td>7, 16, 41, 62, 69, 85</td>
<td>70, 71, 84</td>
<td>29.094</td>
<td>Satisfied</td>
</tr>
</tbody>
</table>

### 4.4 Discrepancies between Estimated and Measured Voltages

As observed in Section 4.3.4, the capability of the proposed method to identify the theft loads is limited to availability of adequate redundant measurements, and that the critical measurements are generally difficult to identify if there is a theft there or not. To the best of our knowledge, there is no solution to identify bad data for critical measurements [13]. At the same time, since the measurement devices in parts of distribution system may not be sufficient for adequate redundancy, the number of identifiable bad data as theft is also limited.
In this section, voltage magnitudes measured by smart meters will be used in the second stage of identification of theft points. After the execution of DSSE accompanied with elimination of bad data, the voltage magnitude of loads is estimated as a part of the system state. In order to extend the capability of the method proposed in Section 4.2, a criterion is proposed to identify critical bad measurements based on the discrepancies between the estimated and measured voltages. The goal of this criterion is to find the acceptable range of discrepancy between the estimated and measured voltages for each load. If the discrepancy between the estimated and measured voltages violates the proposed criterion, then electricity theft at that point may be suspected.

As explained in Section 3.2, the covariance matrix of the estimated state variable, $R_x$, is the inverse of the gain matrix. The covariance matrix, $R_x$, can be partitioned into four submatrices corresponding to the covariance of the estimated voltage angles and voltage magnitudes as

$$
R_x = \begin{bmatrix}
R_x^{\text{angle}} & R_x^{\text{angle,mag}} \\
R_x^{\text{mag,angle}} & R_x^{\text{mag}}
\end{bmatrix},
$$  

(4-11)

where superscripts $\text{angle}$ and $\text{mag}$ refer to angles and magnitudes, respectively. The vector of standard deviation of the estimated voltage magnitudes is the square root of the diagonal elements of the voltage magnitude covariance sub-matrix in (4-12). Therefore,

$$
\sigma_{est} = \sqrt{\text{diag}(R_x^{\text{mag}})}.
$$  

(4-12)
where $\sigma_{est}$ is the vector of the standard deviation of estimated voltage magnitudes. Next, consider the discrepancy between the estimated and measured voltage magnitudes as

$$V_{dis}^{mag} = V_{m}^{mag} - V_{est}^{mag} = \left(V_{actual}^{mag} + e_{m}^{mag}\right) - \left(V_{actual}^{mag} + e_{est}^{mag}\right)$$

(4-13)

where $V_{dis}^{mag}$ is the discrepancy between measured and estimated voltage magnitudes; $V_{actual}^{mag}$ is the actual value of voltage magnitudes with no error; $V_{m}^{mag}$ and $e_{m}^{mag}$ are the measured voltage magnitudes and their measurement errors, respectively; and $V_{est}^{mag}$ and $e_{est}^{mag}$ are the estimated voltage magnitudes and the estimation errors, respectively. Let $\sigma_{m}^{2}$ denote the measurement error variances. Given that $\sigma_{est}$ is calculated in (4-12), the mean and standard deviation of voltage magnitudes discrepancy is

$$E[V_{dis}^{mag}] = E[e_{m}^{mag}] - E[e_{est}^{mag}] = 0$$

(4-14)

$$\sigma_{dis}^{2} = \sigma_{m}^{2} + \sigma_{est}^{2}$$

where $E[V_{dis}^{mag}]$ and $\sigma_{dis}$ are the mean and standard deviation of the discrepancy between the measured and estimated voltage magnitudes, respectively. Thus, the discrepancy between the measured and estimated voltage magnitudes should be within the range of $\left(-3\sigma_{dis}, 3\sigma_{dis}\right)$ with the probability of 99.7% as a property of normally distribute random variables [13]. In the next subsection, the voltage magnitudes discrepancy for the same system of Section 4.3.4 is calculated to find the measured voltage magnitudes which violated the proposed criteria.
4.4.1 Case study

In this case study, it is assumed that the same theft points 7, 16, 41, 62, 69, and 85 are eliminated as a result of multiple bad data identification method. The only theft points left are 70, 71, and 84 with bad data according to SFLA results in Table 4-2. The DSSE is executed for the system of Figure 4-5.

Table 4-3 shows the measured and estimated voltage magnitudes of the system. In this table, the loads 65, 69, 70, 70, 83, 84, and 85 are identified as loads which violated the criteria proposed in Section 4.4. Specifically, the points 65 and 83 are identified as theft loads mistakenly. To explain this miss-identification, one should notice that the proposed method has 3 sources of inaccuracy. The first source is that the parameters of the network have some error, as explained in Section 2.3. Secondly, the off-diagonal elements of the covariance matrix in the approximation in (4-12) are neglected. Thirdly, the proposed method assumes that the voltage magnitude of a load is most sensitive to its active and reactive powers which is less accurate for users at the end of a feeder [133]. The reason that points 65 and 83 are impacted by these inaccuracies is that they are located in very close physical distance from the points with bad data. Thus, their measured and estimated voltage magnitudes are close to those points. Therefore, distinguishing them from theft points is difficult. It should be noted that the proposed method is narrowing down the suspicious points of theft to a very small set of loads which is feasible for utilities to inspect, even when few extra points are identified.
It is important to mention that identification of bad measurements at critical points does not have any impact on the accuracy of DSSE results. The reason is that DSSE cannot be executed after removing a critical bad data and no further estimation is obtainable. However, once the identified meters are inspected and the ones with bad data are fixed, the accuracy of state estimation will be improved.

**Table 4–3 Comparison between Measured and Estimated Voltage Magnitudes Discrepancies with 3σ**

<table>
<thead>
<tr>
<th>Load#</th>
<th>Measured</th>
<th>Estimated</th>
<th>Absolute of Residual</th>
<th>3σ</th>
<th>Load#</th>
<th>Measured</th>
<th>Estimated</th>
<th>Absolute of Residual</th>
<th>3σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0371</td>
<td>1.0371</td>
<td>0.0000</td>
<td>0.0073</td>
<td>44</td>
<td>0.9928</td>
<td>0.9906</td>
<td>0.0022</td>
<td>0.0076</td>
</tr>
<tr>
<td>2</td>
<td>1.0323</td>
<td>1.0376</td>
<td>0.0053</td>
<td>0.0073</td>
<td>45</td>
<td>0.9915</td>
<td>0.9905</td>
<td>0.0010</td>
<td>0.0075</td>
</tr>
<tr>
<td>3</td>
<td>1.0373</td>
<td>1.0373</td>
<td>0.0000</td>
<td>0.0073</td>
<td>46</td>
<td>0.9928</td>
<td>0.9906</td>
<td>0.0022</td>
<td>0.0076</td>
</tr>
<tr>
<td>4</td>
<td>1.0316</td>
<td>1.0308</td>
<td>0.0007</td>
<td>0.0073</td>
<td>47</td>
<td>1.0079</td>
<td>1.0065</td>
<td>0.0014</td>
<td>0.0075</td>
</tr>
<tr>
<td>5</td>
<td>1.0300</td>
<td>1.0292</td>
<td>0.0007</td>
<td>0.0073</td>
<td>48</td>
<td>0.9922</td>
<td>0.9900</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>6</td>
<td>1.0295</td>
<td>1.0287</td>
<td>0.0007</td>
<td>0.0073</td>
<td>49</td>
<td>0.9918</td>
<td>0.9896</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>7</td>
<td>1.0287</td>
<td>1.0280</td>
<td>0.0007</td>
<td>0.0073</td>
<td>50</td>
<td>1.0189</td>
<td>1.0175</td>
<td>0.0014</td>
<td>0.0075</td>
</tr>
<tr>
<td>8</td>
<td>0.9809</td>
<td>0.9803</td>
<td>0.0006</td>
<td>0.0073</td>
<td>51</td>
<td>1.0109</td>
<td>1.0109</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>9</td>
<td>0.9850</td>
<td>0.9847</td>
<td>0.0003</td>
<td>0.0078</td>
<td>52</td>
<td>0.9992</td>
<td>0.9970</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>10</td>
<td>1.0250</td>
<td>1.0247</td>
<td>0.0003</td>
<td>0.0078</td>
<td>53</td>
<td>1.0184</td>
<td>1.0169</td>
<td>0.0014</td>
<td>0.0075</td>
</tr>
<tr>
<td>11</td>
<td>1.0270</td>
<td>1.0212</td>
<td>0.0058</td>
<td>0.0078</td>
<td>54</td>
<td>1.0110</td>
<td>1.0110</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>12</td>
<td>1.0246</td>
<td>1.0242</td>
<td>0.0003</td>
<td>0.0078</td>
<td>55</td>
<td>0.9988</td>
<td>0.9966</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>13</td>
<td>0.9971</td>
<td>0.9913</td>
<td>0.0058</td>
<td>0.0078</td>
<td>56</td>
<td>1.0021</td>
<td>1.0030</td>
<td>0.0009</td>
<td>0.0075</td>
</tr>
<tr>
<td>14</td>
<td>1.0127</td>
<td>1.0127</td>
<td>0.0000</td>
<td>0.0074</td>
<td>57</td>
<td>1.0110</td>
<td>1.0109</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>Load#</td>
<td>Measured</td>
<td>Estimated</td>
<td>Absolute of Residual</td>
<td>$3\sigma$</td>
<td>Load#</td>
<td>Measured</td>
<td>Estimated</td>
<td>Absolute of Residual</td>
<td>$3\sigma$</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------------</td>
<td>---------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------------</td>
<td>---------</td>
</tr>
<tr>
<td>15</td>
<td>1.0235</td>
<td>1.0222</td>
<td>0.0013</td>
<td>0.0074</td>
<td>58</td>
<td>0.9983</td>
<td>0.9961</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>16</td>
<td>1.0233</td>
<td>1.0230</td>
<td>0.0003</td>
<td>0.0079</td>
<td>59</td>
<td>1.0021</td>
<td>1.0030</td>
<td>0.0009</td>
<td>0.0075</td>
</tr>
<tr>
<td>17</td>
<td>1.0296</td>
<td>1.0295</td>
<td>0.0001</td>
<td>0.0074</td>
<td>60</td>
<td>1.0210</td>
<td>1.0209</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>18</td>
<td>1.0181</td>
<td>1.0165</td>
<td>0.0016</td>
<td>0.0075</td>
<td>61</td>
<td>0.9983</td>
<td>0.9961</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>19</td>
<td>1.0293</td>
<td>1.0292</td>
<td>0.0001</td>
<td>0.0074</td>
<td>62</td>
<td>1.0263</td>
<td>1.0271</td>
<td>0.0009</td>
<td>0.0075</td>
</tr>
<tr>
<td>20</td>
<td>1.0218</td>
<td>1.0281</td>
<td>0.0063</td>
<td>0.0079</td>
<td>63</td>
<td>1.0028</td>
<td>1.0068</td>
<td>0.0040</td>
<td>0.0073</td>
</tr>
<tr>
<td>21</td>
<td>1.0247</td>
<td>1.0246</td>
<td>0.0001</td>
<td>0.0075</td>
<td>64</td>
<td>0.9985</td>
<td>0.9979</td>
<td>0.0006</td>
<td>0.0079</td>
</tr>
<tr>
<td>22</td>
<td>1.0101</td>
<td>1.0079</td>
<td>0.0022</td>
<td>0.0076</td>
<td>65</td>
<td>0.9979</td>
<td>1.0059</td>
<td>0.0080</td>
<td>0.0079</td>
</tr>
<tr>
<td>23</td>
<td>1.0095</td>
<td>1.0073</td>
<td>0.0022</td>
<td>0.0076</td>
<td>66</td>
<td>0.9927</td>
<td>0.9927</td>
<td>0.0000</td>
<td>0.0073</td>
</tr>
<tr>
<td>24</td>
<td>1.0093</td>
<td>1.0072</td>
<td>0.0022</td>
<td>0.0076</td>
<td>67</td>
<td>0.9985</td>
<td>0.9979</td>
<td>0.0006</td>
<td>0.0079</td>
</tr>
<tr>
<td>25</td>
<td>1.0093</td>
<td>1.0072</td>
<td>0.0022</td>
<td>0.0076</td>
<td>68</td>
<td>0.9991</td>
<td>0.9954</td>
<td>0.0037</td>
<td>0.0079</td>
</tr>
<tr>
<td>26</td>
<td>1.0089</td>
<td>1.0067</td>
<td>0.0022</td>
<td>0.0076</td>
<td>69</td>
<td>0.9963</td>
<td>1.0047</td>
<td>0.0085</td>
<td>0.0079</td>
</tr>
<tr>
<td>27</td>
<td>1.0085</td>
<td>1.0085</td>
<td>0.0000</td>
<td>0.0078</td>
<td>70</td>
<td>0.9976</td>
<td>1.0068</td>
<td>0.0092</td>
<td>0.0079</td>
</tr>
<tr>
<td>28</td>
<td>1.0068</td>
<td>1.0041</td>
<td>0.0027</td>
<td>0.0078</td>
<td>71</td>
<td>0.9913</td>
<td>1.1003</td>
<td>0.1090</td>
<td>0.0079</td>
</tr>
<tr>
<td>29</td>
<td>1.0046</td>
<td>1.0040</td>
<td>0.0006</td>
<td>0.0078</td>
<td>72</td>
<td>1.0147</td>
<td>1.0146</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>30</td>
<td>1.0003</td>
<td>1.0002</td>
<td>0.0001</td>
<td>0.0076</td>
<td>73</td>
<td>0.9986</td>
<td>0.9964</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>31</td>
<td>1.0027</td>
<td>1.0004</td>
<td>0.0022</td>
<td>0.0077</td>
<td>74</td>
<td>1.0127</td>
<td>1.0128</td>
<td>0.0001</td>
<td>0.0075</td>
</tr>
<tr>
<td>32</td>
<td>1.0089</td>
<td>1.0088</td>
<td>0.0001</td>
<td>0.0076</td>
<td>75</td>
<td>1.0127</td>
<td>1.0127</td>
<td>0.0001</td>
<td>0.0075</td>
</tr>
<tr>
<td>33</td>
<td>1.0027</td>
<td>1.0092</td>
<td>0.0066</td>
<td>0.0075</td>
<td>76</td>
<td>1.0121</td>
<td>1.0129</td>
<td>0.0008</td>
<td>0.0075</td>
</tr>
<tr>
<td>34</td>
<td>1.0002</td>
<td>1.0001</td>
<td>0.0001</td>
<td>0.0076</td>
<td>77</td>
<td>1.0154</td>
<td>1.0154</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>35</td>
<td>1.0015</td>
<td>0.9993</td>
<td>0.0022</td>
<td>0.0077</td>
<td>78</td>
<td>0.9995</td>
<td>0.9973</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>Load#</td>
<td>Measured</td>
<td>Estimated</td>
<td>Absolute of Residual</td>
<td>3σ</td>
<td>Load#</td>
<td>Measured</td>
<td>Estimated</td>
<td>Absolute of Residual</td>
<td>3σ</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-----</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-----</td>
</tr>
<tr>
<td>36</td>
<td>0.9990</td>
<td>0.9967</td>
<td>0.0022</td>
<td>0.0077</td>
<td>79</td>
<td>1.0141</td>
<td>1.0141</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>37</td>
<td>1.0047</td>
<td>1.0032</td>
<td>0.0015</td>
<td>0.0075</td>
<td>80</td>
<td>1.0099</td>
<td>1.0115</td>
<td>0.0016</td>
<td>0.0075</td>
</tr>
<tr>
<td>38</td>
<td>1.0205</td>
<td>1.0204</td>
<td>0.0001</td>
<td>0.0076</td>
<td>81</td>
<td>1.0167</td>
<td>1.0166</td>
<td>0.0001</td>
<td>0.0076</td>
</tr>
<tr>
<td>39</td>
<td>1.0004</td>
<td>0.9982</td>
<td>0.0022</td>
<td>0.0077</td>
<td>82</td>
<td>1.0004</td>
<td>0.9982</td>
<td>0.0022</td>
<td>0.0077</td>
</tr>
<tr>
<td>40</td>
<td>1.0011</td>
<td>1.0017</td>
<td>0.0006</td>
<td>0.0075</td>
<td>83</td>
<td>1.0011</td>
<td>1.0090</td>
<td>0.0079</td>
<td>0.0077</td>
</tr>
<tr>
<td>41</td>
<td>1.0208</td>
<td>1.0207</td>
<td>0.0001</td>
<td>0.0076</td>
<td>84</td>
<td>1.0045</td>
<td>1.0123</td>
<td>0.0078</td>
<td>0.0075</td>
</tr>
<tr>
<td>42</td>
<td>0.9995</td>
<td>0.9973</td>
<td>0.0022</td>
<td>0.0077</td>
<td>85</td>
<td>1.0039</td>
<td>1.0123</td>
<td>0.0083</td>
<td>0.0076</td>
</tr>
<tr>
<td>43</td>
<td>1.0199</td>
<td>1.0185</td>
<td>0.0014</td>
<td>0.0075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proposed method has been tested on various distributions of theft points throughout the network and similar performance of the proposed case study has been observed. In practice, it is likely that utilities would have an assessment of suspicious regions for theft in their distribution systems. Therefore, the theft points location would be system dependent.
CHAPTER 5: SUMMARY OF CONTRIBUTIONS AND FUTURE WORKS

5.1 Conclusions and Contributions

The main objective of this thesis was to address challenges in using AMI in order to obtain an accurate distribution system state estimation, and utilize its results to enhance the transmission system state estimation. Furthermore, as an application of DSSE, electricity theft identification method was also proposed. Based on the contributions presented in this thesis, all three original objectives have been achieved as summarized below.

Objective 1

In the first objective, the challenge of using AMI with asynchronous measurements has been solved by proposing a method based on modeling load variations during any given interval of the day. In Chapter 2, the total error of a power measurement signal is assumed to consist of two terms: i) the measurement error due to imprecision of the meter device, which is modeled as a normally distributed random variable; and ii) the load variation between the measurement time and DSSE execution. A statistical analysis on load variations has been conducted. According to this analysis, it is shown that the load variation can be represented as a random variable with normal distribution, and its error increases as the time passes from the sampling instance and measurement becomes out-of-date. The DSSE formulation has been modified to incorporate the proposed approach of
appropriately modifying the variances in the estimation procedure as to give more weight to the fresh measurements and reduce the weight of the old measurements, accordingly. This modification and methodology is compatible with commercial packages, which makes it implementable in industry.

It is explained that by using smart meters, the distribution system would be observable and DSSE can be successfully executed. The proposed DSSE was tested against traditional DSSE on IEEE 13-bus and IEEE 123-bus distribution systems assuming that smart meters' measurements are updated every 15 minutes. One of the criteria was that the developed method should be applicable for both radial and meshed networks. Therefore, IEEE 123-bus system was modified to include loops. According to the simulation results, the proposed method is significantly better than the traditional method and improved the accuracy of DSSE by 14% to 16% for the considered IEEE test systems. Furthermore, the robustness of this method is verified against highly fluctuating unpredictable load profiles. Moreover, as the noise generated is a random phenomenon, the simulation is repeated for 10000 times and the robustness of the proposed method and its consistency in improving the accuracy of the estimation is demonstrated.

**Objective 2**

The second objective of this thesis was to enhance the transmission state estimation using AMI that is becoming available at the distribution level. The proposed TSSE method is based on incorporating the results of DSSE into the measurement set of the transmission system. Specifically, PQ bus voltage magnitudes and active and reactive powers estimated
by DSSE at the feeder heads are incorporated into TSSE in the form of measurements at the corresponding PQ buses. The weights of these additional measurements are calculated from the gain matrix in DSSE as a part of the method proposed in Chapter 3. Four areas of improvements on two benchmark systems based on the IEEE 14-bus and IEEE 57-bus transmission systems that are connected to the IEEE 13-bus distribution system at their PQ buses have been investigated.

Firstly, it is shown that an unobservable transmission system (due to lack of measurements, or measurement and communication failures) can become observable using the additional DSSE results according to the proposed approach. In particular, 2 scenarios are presented in which a critical measurement is not available and the IEEE 14-bus transmission system is unobservable. Then, for each scenario, it is explained that the proposed method provides adequate measurements which restore the observability of the system.

Secondly, as TSSE typically suffers from ill-conditioning, the condition numbers of the two benchmarks are presented in a simulation. The results show that the proposed TSSE method improved the condition number by 2 orders of magnitude in comparison to the tradition TSSE method. Moreover, the number of iterations which takes TSSE to achieve convergence is compared between the proposed TSSE and the traditional TSSE. It is showed that the proposed TSSE converges with 60% faster convergence rate. Therefore, as the proposed TSSE expands the size of measurements by approximate 60%, it is concluded
that the computational burden imposed by the proposed approach is almost compensated by its convergence rate.

Since the proposed TSSE increases the redundancy of the system, it facilitates the bad data identification methods. In a comparison to traditional TSSE, in a case where traditional TSSE fails to identify a bad datum, it is demonstrated that the proposed TSSE method increases the capability of the normalized residual based bad data identification method and identifies the bad datum.

The proposed TSSE resulted in better numerical conditioning, higher rate of successful bad data identification and its subsequent elimination, accurate DSSE calculation due to advance metering infrastructure, all of which positively contributes to the improved resulted accuracy of the proposed TSSE. However, it is explained that improvement of accuracy also depends on the resolution of smart meters measurements. Simulation results showed that when the resolution of smart meters measurements decreases, the improvement of accuracy of the proposed TSSE decreases too.

Objective 3

The third objective was to propose a methodology to identify the theft points in distribution systems. In Chapter 4, the idea of modeling electricity theft as bad data is presented. A two stage method is proposed to identify theft points in a system. In the first stage, assuming adequate redundancy of measurements, multiple bad data identification problem is using a heuristic optimization method is proposed to identify theft points. In the
second stage, in order to identify the theft points without redundant measurements, a criterion is designed based on the discrepancy of measured and estimated voltage magnitudes.

The results of the proposed method are demonstrated on the IEEE 123-bus distribution system. It is shown that in the first stage, some of the theft points are identified. However, theft points without redundant measurements are missed. In the second stage, the proposed method identifies the theft point without redundant measurements. It is explained that, in the process of identifying theft points, the proposed method may identify some points mistakenly as theft points. However, it is discussed that the identification of suspicious smart meters, even combined with some misidentifications, is beneficial since it is feasible for utilities to inspect a number of locations which is small in comparison to the entire distribution system.

5.2 Potential Impacts of Contributions

The problem of non-synchronized AMIs measurements in some cases has discouraged utilities from investing on DSSE. Without DSSE, most of the applications in distribution systems are not implementable. For instance, real-time power quality measures such as Voltage/Var optimization (VVO) are not practiced (or are poorly practiced) due to lack of monitoring mechanisms in place.

The proposed DSSE is reliable and can be relatively easy implemented in industrial programs, and therefore may encourage utilities to use non-synchronized AMI
measurements in order to estimate the state of the distribution system. This would not be an expensive investment for utilities since the proposed DSSE is compatible with existing commercial packages. As a result, a better monitoring of the system would be achieved. This is a fundamental step in controlling distribution systems and practice advanced applications in DMS.

Presently, the design of smart meters has been mainly focused on dynamic billing applications. As smart meters may be used in real-time operation of power systems, in the future, the manufacturers may redesign smart meters to provide additional metering functions.

Operation of transmission systems at lower voltage level often suffers from unobservability. The proposed TSSE method renders these systems observable. This would enable the operators to have a better insight about the system state. Also, real-time applications such as optimal power flow would include a larger portion of the system and result in more comprehensive solutions. Also, improvements in numerical conditioning and accuracy of the system would have a direct impact on the accuracy of important applications such as transient and voltage stability analysis, since these applications are based on TSSE results. Therefore, the performance of real-time applications would be improved, and the control decisions are less likely to be problematic.

Finally, the proposed theft identification method based on DSSE results is a novel attempt at identifying the manipulated measurements. Electricity theft has greatly impacted the revenue of utilities, and the proposed method has a potential to reduce this damage. Also,
sometime smart meters may expand the vulnerability of metering infrastructure, and because of that utilities may be disincentivized from deployment of more smart meters. Consequently, the implementation of applications such as dynamic pricing and demand response management, which require smart meters, may be hindered. However, the proposed method will be a positive influence towards installing more smart meters by reducing the risk to energy theft.

5.3 Future Work

One possible improvement in the DSSE method proposed in Chapter 2 could be considering various statistical distributions and finding a model which represents the load more accurately than the normal distribution based models. In the next step, combining the chosen distribution with normally distributed measurement noise should yield more accuracy. Moreover, trying other state estimation techniques such as nonlinear Kalman Filter could result in improvements of the results. Another area of research could be studying the impact of modeling loads as dependent random variables which may conform to similar profiles.

As mentioned in Chapter 3, if the resolution of smart meters’ measurements is very low, the DSSE results will not be accurate enough to improve the accuracy of TSSE when fast load variation occurs. To find a method that may mitigate this problem could be a research topic. For example, developing accurate load forecasting methods and incorporating them into the SE framework could improve the results. The reliability of the developed forecasting methods should be studied. It is important to notice that in real-time
applications, the reliability of the methods used is more important than it is in off-line applications such as planning.

The proposed theft identification method may result in mistakes and classify some good measurements as manipulated ones. Designing a more accurate criterion for identifying non-redundant bad measurements would improve the results. Combining some developed methods such as game theory with the proposed method could contribute to better theft identification solutions. It is important to notice that the proposed methods should not use the load behavior profile since this kind of information is recognized as a violation of private information of customers.
REFERENCES


[42] [online] Available at: http://hdl.handle.net/10986/20786 [May. 1, 2013].


[91] [online] Available at: http://utilsublbs.ubc.ca/ion/ [Aug. 1, 2013].


[105] [online] Available at: http://www.ee.washington.edu/research/pstca/ [May. 1, 2015].


[115] P. Venkatesh, R. Gnanadass, and N. P. Padhy, "Comparison and application of evolutionary programming techniques to combined economic emission dispatch with


