4-12-2022 11:00 AM

Multi-Device Data Analysis for Fault Localization in Electrical Distribution Grids

Jacob D L Hunte, *The University of Western Ontario*

Supervisor: Lutfiyya, Hanan, *The University of Western Ontario*
Co-Supervisor: Haque, Anwar, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Computer Science

Abstract

The work presented in this dissertation represents work which addresses some of the main challenges of fault localization methods in electrical distribution grids. The methods developed largely assume access to sophisticated data sources that may not be available and that any data sets recorded by devices are synchronized. These issues have created a barrier to the adoption of many solutions by industry. The goal of the research presented in this dissertation is to address these challenges through the development of three elements. These elements are a synchronization protocol, a fault localization technique, and a sensor placement algorithm.

The synchronization protocol addresses the dependency on synchronized data by allowing the devices themselves to synchronize their data. This is accomplished by establishing relationships between temporal events, transient signals and the devices recording these events and signals. The protocol establishes a relationship between transient signals emitted from standard equipment in electrical grids and the fault event which is then leveraged to synchronize the recorded fault data. The method has shown very promising results in synchronizing data from multiple devices.

The fault localization technique determines the location of a fault in a complex distribution grid. This is achieved by establishing spatial-temporal relationships between transient signals and distances in graphical representations of the distribution grid. This method provides precise fault locations in challenging electrical distribution grid structures. It also allows areas of the distribution grid to be classified based on how well they are monitored. This classification component predicts how well a distribution grid is monitored by the devices used.

The sensor placement algorithm leverages the classification component established in the fault localization technique to optimize the placement of the fault localization devices in the distribution grid. Here, an efficient algorithm is created for determining the best location for a given number of fault localization devices.

The combination of the synchronization protocol, a fault localization technique and a sensor placement algorithm and their minimal data requirements constitutes a uniquely complete

solution that may set it apart from existing work and make it more attractive to the industry for adoption in electrical distribution grids.

Keywords

Electrical Grids, Smart Grids, Distribution Grids, Synchronization, Fault Localization, Optimization, Data Synchronization, Multiple Sensors, Distributed Systems.

Summary for Lay Audience

Fault localization in distribution grids has become a key area in the development of Smart Grid technology. According to a report made for the US Department of energy the annual cost of power interruptions to customers are in the billions of US dollars. A fault is an event that causes the grid to operate incorrectly resulting in an abnormal and often dangerous flow of power (like a fallen power cable or tree falling on a power cable). Fault localization is the process of determining the location of the source of the fault.

There are two areas to consider when performing fault localization in electrical grids. Fault localization can occur in the transmission grid and the distribution grid. Fault localization in the distribution grid is significantly more difficult that in the transmission grid. Fault localization solutions for the transmission grid have been quite successful. As such, there has been a wide range of methods developed to enhance fault localization methods used in the transmission grid so that these methods can be applied to the more complex distribution grid. The work has had varying levels of success, but these solutions are generally not adopted by the industry. This is because the methods developed largely assume access to sophisticated data sources that may not be available and that any data sets recorded by devices are synchronized.

The work presented in this dissertation seeks to address the current issues and provides a more efficient fault localization technique in relation to the data needed. The method is a multi-device localization method that uses a minimal amount of readily available data while maintaining a high level of accuracy. This thesis also provides an effective method of synchronizing all of the data and devices used for the localization method presented and establishes a reliable way to determine the best locations for those devices in the electrical grid.

Co-Authorship Statement

This thesis includes three articles written by the author, Jacob Hunte and supervised by Dr Hanan Lutfiyya and Dr Anwar Haque. Two of the articles have been published. The articles represent research conducted at the University of Western Ontario during the fulfillment of the author's doctoral research. These works are presented in chapters three, four and five:

Chapter 3: Jacob Hunte, Dr. Hanan Lutfiyya and Dr. Anwar Haque. Device Synchronization for Fault Localization in Electrical Distribution Grids

I am the principal author of this peer reviewed conference paper which was part of the IEEE 7th World Forum on Internet of Things in June 2021. I developed the synchronization method, wrote the manuscript explaining the process and presented the work at the conference. Dr. Hanan Lutfiyya and Dr. Anwar Haque acted in a supervisory role and provided recommendations for presenting the content of the work and editing and reviewing the paper.

Chapter 4: Jacob Hunte, Dr. Hanan Lutfiyya and Dr. Anwar Haque. Fault Localization in Smart Grids Using Segmentation

I am the principal author of this peer reviewed conference paper which was part of the IEEE International Conference on Communications in June 2021. I developed the fault localization method, wrote the manuscript explaining the process and presented the work at the conference. Dr. Hanan Lutfiyya and Dr. Anwar Haque acted in a supervisory role and provided recommendations for presenting the content of the work and editing and reviewing the paper.

Acknowledgments

I would first like to express my gratitude and appreciation for my supervisors Dr. Hanan Lutfiyya and Dr. Anwar Haque for their continued support and guidance during the course of my doctoral studies. They provided an exceptional environment to facilitate my personal and professional development during my studies. Dr. Hanan Lutfiyya and Dr. Anwar Haque provided an extraordinary project and avenue of research which allowed me to work very closely with industry. Our meetings were always an exciting discussion about the project that fueled its growth and encouraged me to push its development further. My supervisors' involvement was key to not only partnering with industry on an exciting project but also converting the project into a potentially licensable work.

I thank the industry partners that provided the initial challenge and provided the necessary information and data about the industry and their support of its continued development as a functional solution. I also acknowledge Dr. Tamer Mohamed and Patrick Therrien and the team at World Discovery and Dr. Jai Chatterjee for their support in the project. Patrick Therrien, Dr. Tamer Mohamed and Dr. Jai Chatterjee greatly assisted in refining the parameters of the project to align with the requirements of the patenting process and in the filling of the provisional patent.

I also thank the contributions of the anonymous reviewers of my publications whose feedback helped improve the quality of my manuscripts. I thank Dr. Marc Moreno Maza, Dr. Dan Lizotte and Dr, Kostas Kontogiannis for very enlightening courses and access to resources that provided very insightful views on concepts that associated with my research. My sincere thanks to my TSP, supervisory and examination committees for their valuable recommendations and comments and for the time taken for discussions and reading my work. I also thank the faculty and staff of the Western University Department of Computer Science for all of their support and instruction during the course of my studies.

My sincere thanks and acknowledgement to my family, whom without I would not be here. Their support over the years has been a source of motivation and comfort during all my studies. I also thank my friends here in Canada who helped me transition into life here and my friends in Barbados who gave me the support despite being very far away. These friendships, both new and old, allowed me to establish the balance necessary between work and rest that was the crucial to all of my achievements thus far, and will continue to be in the future. I am most grateful to you all.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Introduction

1.1 Motivation

The National Institute of Standards and Technology reports of the Smart Grid Interoperability Standards Roadmaps are part of a mandate in the Energy Independence and Security Act (EISA) [Arnold et al. 2010], [Gopstein et al. 2021]. These reports highlight the need for further development of infrastructure for the Smart Grid as recommended by industry experts that are part of the National Institute of Standards and Technology (NIST) Domain Expert Working Groups (DEWGS).

The NIST reports set Smart Grid development as a pressing issue. They also list major challenges associated with this goal. Among the list provided were challenges in the areas of:

- The complexity of the Smart Grid requiring instantaneous, automated processes.
- Smart equipment with local intelligence that can carry out instructions when remote analysis is unnecessary or not economical.
- Data management to ensure accuracy, time-stamping and consistency across data sources.
- Application development (programs, algorithms, calculations and data analytics) to solve increasingly complex problems with accurate and timely data to deliver quicker and more accurate results.

Discussions with industry experts in smart grid technology development and electrical power distribution companies [Industry-Partner 2018] allowed a focus area to be identified. Industry professionals [Industry-Partner 2018] highlighted fault localization and grid monitoring as a pressing need. This is because a reduction in power outage times and detailed power consumption data is considered crucial and fault localization and grid monitoring are key components to achieving this. According to a report made for the US Department of energy the annual cost of power interruptions to customers are over 20 billion US dollars [LaCommare and Eto 2004]. In 2003 a power outage caused over 50 million customers to go without power [LaCommare and Eto 2004]. This provided a narrow focus point within the broader smart grid development area. There has been work done in smart grids and fault localization. However, there is still a pressing need for more accurate and cost-effective fault localization methods to be developed as stated in the Smart Grid Interoperability Standards Roadmaps [Arnold et al. 2010], [Gopstein et al. 2021]. In addition to improving the accuracy of solutions, fault localization methods have been focused on the high voltage transmission grid (power transmission across large areas like provinces and countries). Many electrical power distribution companies are responsible for localized power distribution (power distribution within a city or small area). As such, the development of fault localization methods within the distribution section of the grid is of utmost importance to these companies. The distribution grid has proven to be much more challenging for fault localization [Jun et al. 1997]. The development of a method specifically for the distribution grid, with a focus on cost effectiveness and accurate fault localization, would allow power distribution companies to align the functionality of their distribution grids with that of a smart grid. Within the research, there has been much work done in developing fault localization techniques for the distribution grid [Takagi et al. 1982], [Girgis et al. 1993], [Myeon-Song et al. 2004], and [Magnago and Abur 1998]. There has been reasonably successful improvements on the pre-existing methods by applying new analytic methods [Pourahmadi-Nakhli and Safavi 2011], [Wen and Chang 1998, Srinivasan et al. 2000]. The work done in the use of multiple devices in the distribution grid by [Jun et al. 1997] has been very successful. Within this dissertation the term device and sensor refer to equipment that records electrical data and are used interchangeably. The use of multiple devices has shown much promise and has resulted in significant improvements in distribution grid fault localization. The methods developed largely assume access to sophisticated data sources that may not be available, frequent transfer of data between devices and that any data recorded by devices are synchronized. The result of this is that these solutions are generally not adopted by the industry. The next step is to further develop the initial multi-device concept to create new methods for fault localization specifically for the distribution grid. These methods should use a minimal amount of data that is widely available and address the challenge of synchronization with the use of multiple devices.

1.2 Background

1.2.1 Electrical Grids

Power is a critical part of every industry. The average power consumed yearly has consistently increased [Prajapat 2018]. The industry has started to adjust the sources of power over time to match the steady increase in demand through the use of renewable energy sources like wind and solar power. Alternative sources of energy are examined by [Hoffert et al. 2002] and provide a concise view of different types of energy production. These alternate sources of energy mitigate an increase in demand for traditional sources of power like coal and oil. However, the infrastructure that handles the transport of this power (the electrical grid) remains a primary and critical component in the industry as the increase in demand for power must travel over this network.

The work in [Pabla 2005] explains how the electrical grid works in great detail and highlights the many components that make up the electrical grid and details on their usage. Figure 1.1 is a summarized view of the main elements of the electrical grid. There are three main sections where each segment is responsible for a major role in delivering power to consumers. These are power generation, transmission, and distribution.

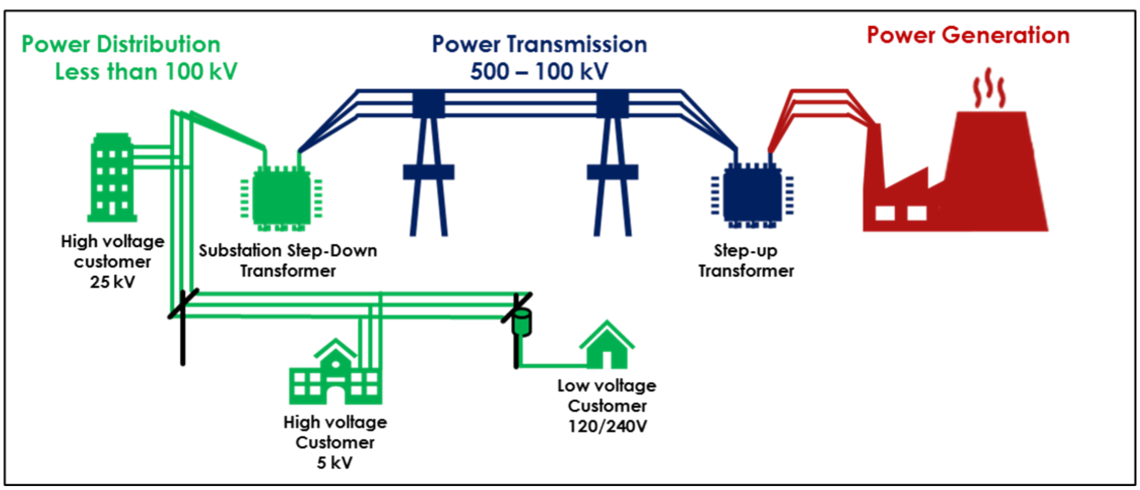

Figure 1.1. Basic structure of electrical grid

1.2.2 Power Generation

Power is generated traditionally at power plants located far from the cities that would be using the generated power. Power is generated as alternating current. The magnetic induction causes current to flow relative to the polarity of the end of the magnet. If the end of the magnet next to the coil is positive (red), then the current flows towards the magnet. If it is the negative end, then the current will flow away from the magnet. As the magnet rotates, and the positive end gets further away from the coil the current weakens until it goes to zero. As the negative end of the magnet moves closer to the coil the current increases again but negatively (negative and positive are used in this context to indicate direction and do not cancel each other out). The current increases until it reaches an upper limit which is represented by the magnet pointing directly at the coil. Figure 1.2 shows the rotation of a magnet near a coil with a closed circuit resulting in induction. It also shows the current alternating from its maximum positive and negative values at each stage resulting in a sine wave.

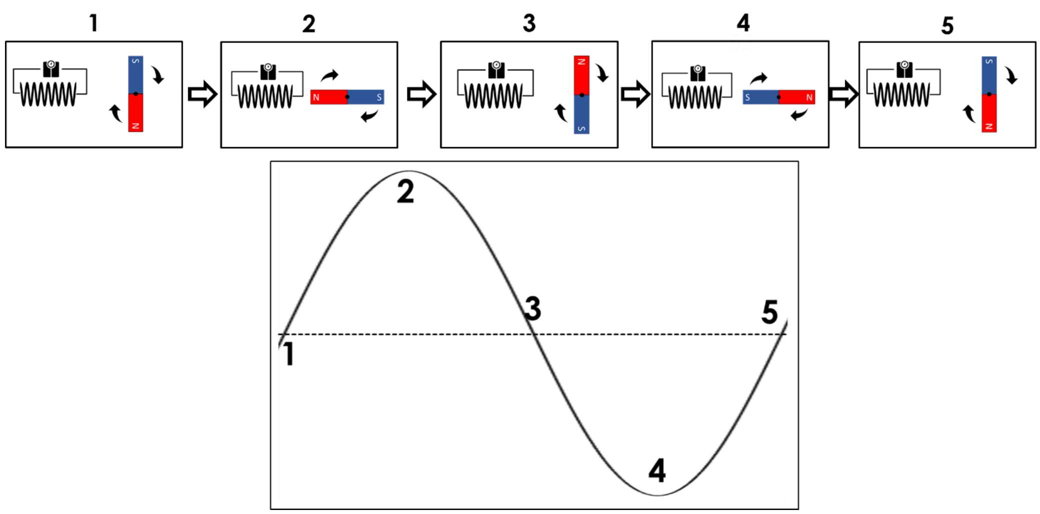

Figure 1.2. Generation of power from induction

This effectively creates an alternating current between the maximum current in both directions. When this is plotted over time it creates a sine waveform as seen in figure 1.2. Power is generated in three phases as seen in figure 1.3.

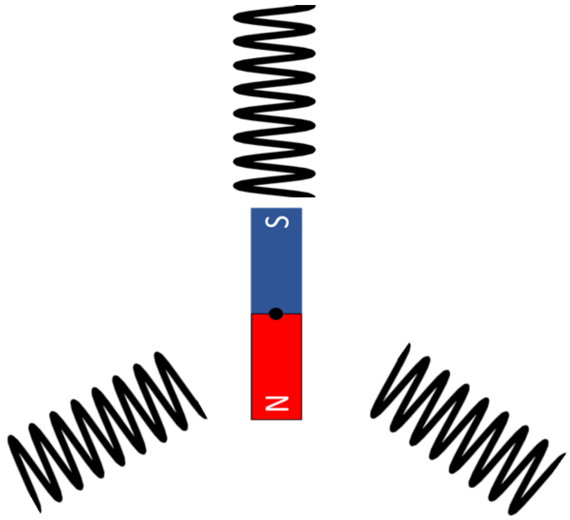

Figure 1.3. Three phase power generation

Each coil is separated by 120 degrees and generates its own alternating current as the magnet rotates. Each coil would therefore have power but at different times. Each coil provides a *phase*. If the current on each coil is plotted over time and placed on the same graph the sine waves show what the three-phase alternating current looks like with respect to the change in current over time. This is shown in figure 1.4.

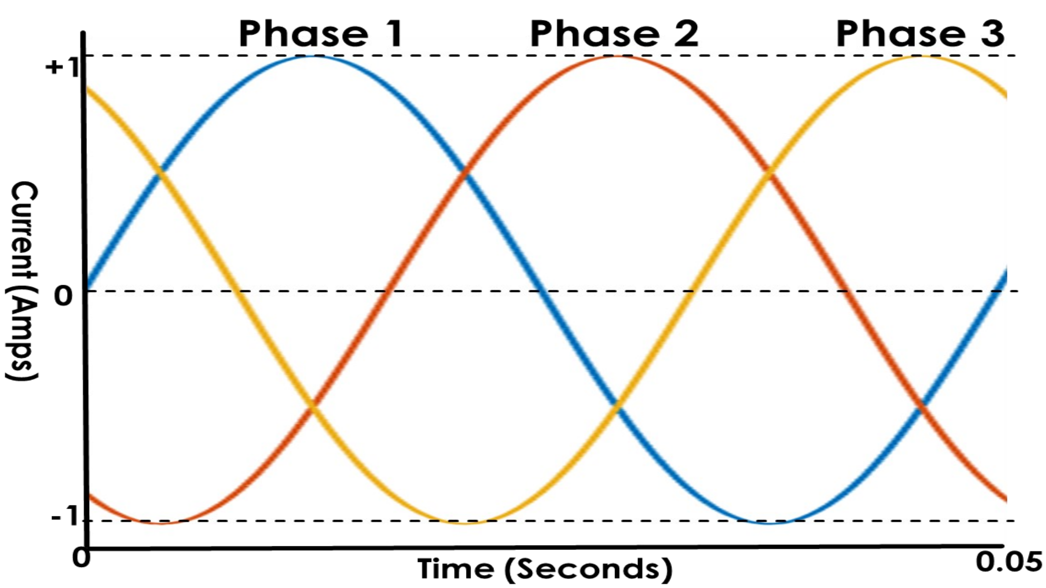

Figure 1.4. Three phase sine waves

1.2.3 Power Transmission and Distribution

After power is generated, it must then be transported to the consumer. This process starts at the transmission substation near the power plants which step up the voltage to between 100 and 500kV [Industry-Partner 2018]. This allows it to be sent along the transmission lines with little loss of power. Some industrial sized manufacturing plants may be supplied with power at this stage. Most of the power continues along the transmission lines to the cites or towns. Once it reaches the cities and towns it is then passed through a distribution substation. This substation steps the power down to anywhere between 60 and 4 kV. At this point the power is usually transported on overhead lines seen within cities. Some large businesses and compounds would usually be supplied at this stage. After this point, the power reaches neighborhoods and smaller businesses. It is fed into transformers which step down the voltage to the final voltage used in houses etc. 125 or 250V. Some houses in cites are powered using underground cables. Rather than having a transformer on the pole and having it overhead, the transformer is placed on the ground and the stepped down lines are run underground to the houses.

With increasing demand being placed on the grid as power needs grow globally the need to improve and ensure the continuity, efficiency and reliability of the electrical grid becomes increasingly important. There are several faults that can occur at different stages throughout the grid. These faults, given the increasing amount of power passing through the grid, can cause massive damage to the equipment and even cause harm to people [Prajapat 2018].

1.2.4 Faults

In the electrical grid power is usually distributed using three phases. This means that there are three lines used to move power from one location to another. With reference to the three-phase system of power, there are typically two main types of faults. These are open circuit and short circuit faults.

1.2.4.1 Open Circuit

Open circuit faults are caused by a break in the conductors. These conductors are the lines or cables used to transport electricity through the grid. The material in conductors allows the electric current to flow freely. A break in these conductors disconnects part of the circuit from the power source preventing power from reaching its destination. In an electrical grid, such a break can occur in any or all the three phases causing single-phase, two-phase or three-phase breaks. These are graphically depicted in figure 1.5.

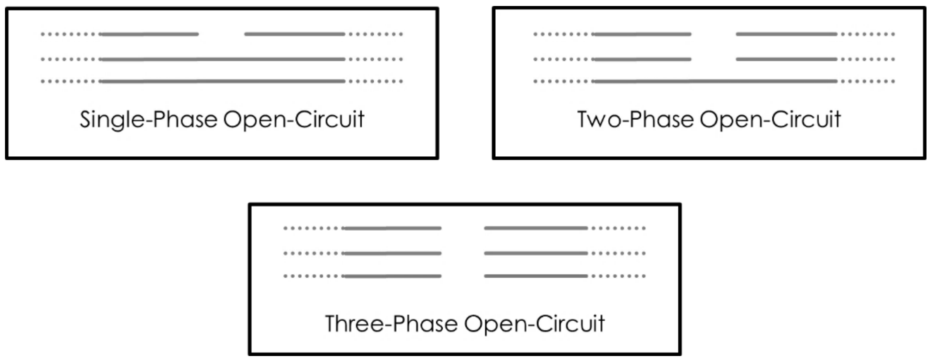

Figure 1.5. Open Circuit Faults

1.2.4.2 Short Circuit

This type of fault occurs when there is a connection between a power cable and the ground or two or more adjacent power cables which causes an excessive amount of current to flow through the electrical grid. In the electrical grid these short circuits can occur along any of the three phases used to transport power in the distribution grid and may or may not be grounded, resulting in the six short circuit faults listed below and shown in figure 1.6.

- Three-phase – a connection between each of the three phases only
- Three-phase to ground – a connection between all three phases and the ground
- Phase-to-phase – a connection between two phases only
- Single-phase to ground – a connection between one phase and the ground
- Two-phase to ground – a connection between two phases and the ground
- Phase to phase plus single-phase to ground – a connection between two phases and having the third phase connected to ground

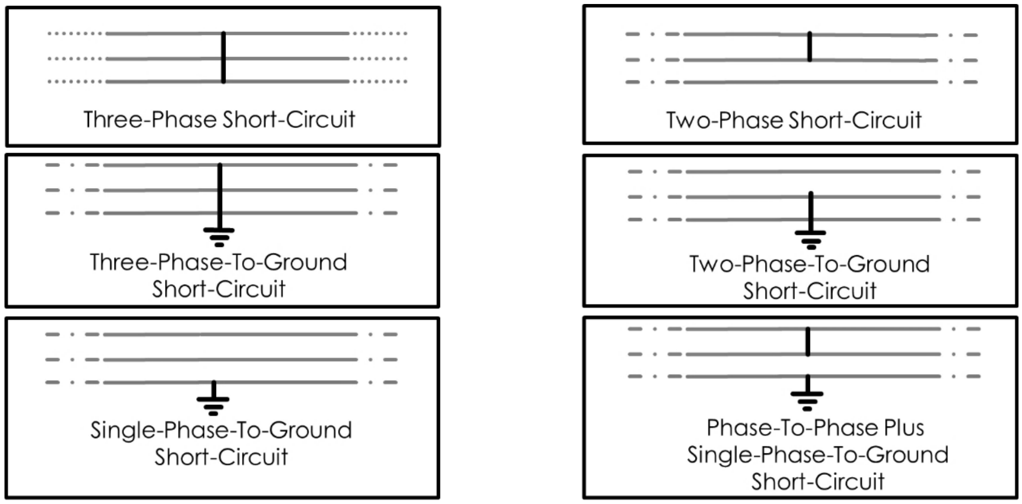

Figure 1.6. Short Circuit Faults

1.2.4.3 Symmetrical and Unsymmetrical

With a three-phase system there is the notion of the system being balanced. This means that the power generated is distributed evenly along each phase. When one of the above faults occurs, it could cause the system to be unbalanced. Unsymmetrical faults cause the system to become unbalanced and symmetrical faults allow the system to remain balanced. Almost all the faults that occur in a distribution grid are unbalanced short circuit faults. These make up 60 to 90 percent of faults [Prajapat 2018]. These faults are those that do not involve all three phases like a two phase to earth fault or a two-phase open circuit. Very few of the faults that occur would be balanced faults which simultaneously involve all three phases like the three-phase open circuit or three-phase to earth faults.

1.2.4.4 Fault Protection Devices

Power companies constantly seek ways to mitigate damages and increase the continuity, efficiency, and reliability of the power grid. The industry has and continues to develop a number of methods to achieve this. Protection devices are traditionally isolated devices that function on their own. As such they would not provide automated digital detection or detailed localization capabilities for the system. The following are examples of some initial devices used to protect the electrical grid: relays, fuses, reclosers, breakers.

Faults can occur at different stages throughout the electrical grid. These faults, given the amount of power passing through the grid, can cause massive damage to the equipment and even cause harm to people. Should a fault occur, the grid is designed to quickly stop sending power to that section of the grid. This protects people, the electrical grid, and its equipment. Specialized equipment like reclosers, relays, fuses and switches allow the power to be shut off quickly [Prajapat 2018]. However, the process of restoring the power is not as efficient.

1.2.4.5 Handling Faults

When a fault occurs, the actual process that follows to resume normal operation of the grid is given by the following five steps:

1. The fault is detected by reclosers, relays, fuses, breakers etc.
2. The power to that area is turned off usually automatically by the reclosers and relays.
3. The fault is located which is very time consuming.
4. The fault is fixed.
5. Power is restored to the area.

Of these five steps the most critical to the process are detecting the fault and localizing the fault. The detection devices that detect the fault and automatically turn of the power in that area traditionally have no way of reporting back to the company that there has been a fault. Most companies traditionally rely on customer call-ins to notify them of faults [Industry-Partner 2018]. If it is not known that a fault has occurred, then the process of handling the fault cannot even be started. As mentioned, the fault detection devices are able to automatically turn off power in the area. The next step after turning of the power is to find the fault. This step is quite complex and difficult to perform. Using call-ins or even in a scenario where the relay can indicate that it has detected a fault still leaves a large area in which the fault could have occurred as shown in figure 1.7.

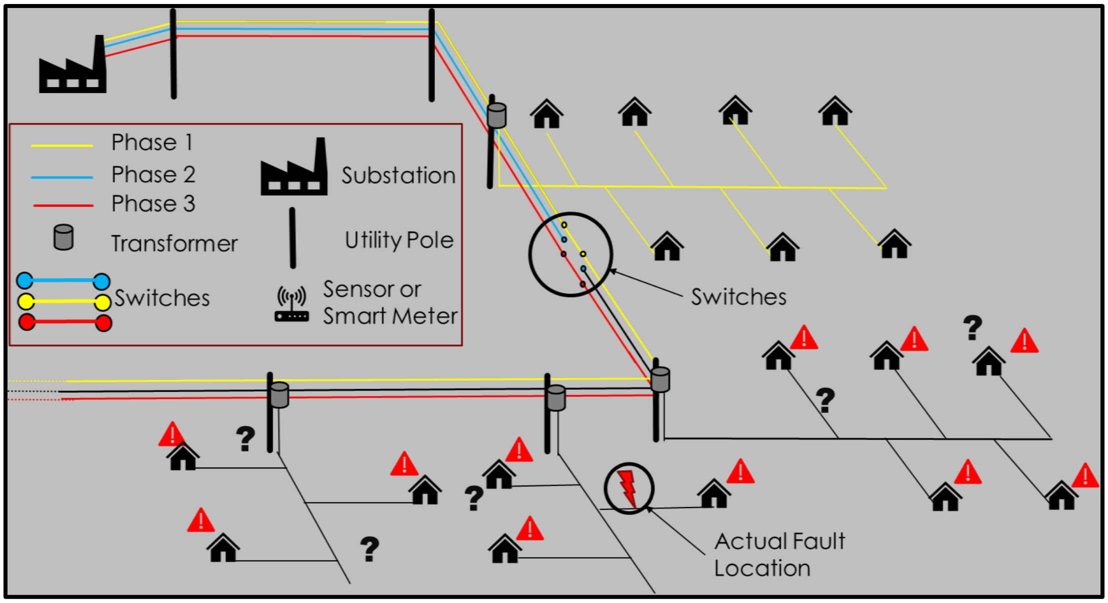

Figure 1.7. Example of impact area for a fault

In figure 1.7, a fault occurs in the area powered by phase 2. When the fault is detected by the switch/relay it opens and shuts of power to the entire area powered by that phase. There is actually no knowledge of the location of the fault in the area. It is only known that something has gone wrong and caused the power to be turned off. If this area covers a two- or four-mile area, that entire area has to be visually inspected. This challenge is compounded by the fact that some of these areas of the grid are buried underground. Recent developments in fault localization by [Sapountzoglou et al. 2019], [Gabr et al. 2017], [Lazaropoulos 2017] have sought to address the need for a modern localization approach with varying levels of success. These approaches are discussed further in Chapter 2.

1.3 Goals and Contributions

1.3.1 Goals

The main goal of the research contained in this dissertation is to develop a fault localization method that aligns with the needs of power distribution companies and can provide movement towards Smart Grid development as outlined in the NIST Smart Grid Interoperability Standards Roadmaps [Arnold et al. 2010] and [Gopstein et al. 2021]. This method would extend the existing work in fault localization by further exploring multi-device fault localization. The solution is developed to localize the most common types of faults occurring on the distribution grid. The solution therefore localizes any low impedance fault that occurs on the grid. Low impedance faults include unbalanced short circuit faults that make up 60% to 90% of all faults on the distribution grid and the rarer balanced short circuit faults. Together, low impedance faults (short circuit faults) account for up to 95% of faults on the electrical grid [Prajapat 2018]. Specifically, the goal is to develop a new method that uses multiple devices in order to address the challenges of low impedance fault localization in the more complex distribution grids (discussed in Chapter 2). This method will also incorporate data synchronization which is seldom considered in fault localization solutions. The method will also optimize the locations of the devices in the grid. Under the main goal three focus areas for development are therefore identified:

Synchronization of Fault Localization Devices (Chapter 3): One area of focus is the exploration of unique ways of establishing relationships between temporal events, transient signals and the devices recording these events and signals. This is done in an effort to provide an alternate technique for synchronizing the data captured across multiple devices that may be present in the electrical grid.

Fault Localization Using Multiple Devices (Chapter 4): Another focus will be to establish spatial-temporal relationships between transient signals and fault distances in graphical representations of the distribution grid. This would allow the characteristics of the distribution grid to be considered in the development of a new solution for localizing faults on the distribution grid.

Optimization of Fault Localization Device Placement (Chapter 5): The final focus area will be to create an optimization method for determining the most effective location for the devices used in the new fault localization technique. This method must maximize the use of each device to incorporate efficiency into the overall solution. The optimization of device placement would allow a reduction in implementation costs of the overall solution and provide a means for the industry to perform cost benefit analysis on device utilization quantities.

1.3.2 Contributions

The need for new technology and more research to be applied in the electrical grid to improve the existing infrastructure has been made from both the government and industry. This need is officially expressed in the Smart Grid Interoperability Standards Roadmaps [Arnold et al. 2010] and [Gopstein et al. 2021]. These documents outline a plan for the power industry to improve the existing systems and infrastructure used for power distribution. This shows that there is an intention to move the industry towards a smarter grid that is gaining momentum and is present in both government and industry. The process of achieving this is complex as it involves development in many areas like policies, infrastructure, procedures, and technology. In addition to the attention of congress in the area, the power industry itself and researchers have started developing technology and methods to help move the power industry forward. Following this movement, research has started developing crucial areas of power distribution. Research in fault localization in distribution grids has sought to apply existing fault localization techniques used in the transmission grid to the distribution grid. Though there has been some success in simulated tests, the existing solutions result in multiple fault locations being calculated due to the increased complexity of the distribution grid (discussed in Chapter 2.2). The distribution grid has presented challenges in developing solutions for fault localization. The most promising approaches show that the use of multiple grid-monitoring devices may be a key element in addressing those challenges (Chapter 2.3). There is a need to further explore the potential of using multiple devices to address the challenges of the distribution grid. The use of multiple devices requires some form of synchronization. There is a general assumption that the clocks on all devices are synchronized using a third-party synchronization scheme like GPS used in [Gopakumar et al. 2015]. There is at present a lack of fault localization methods that do not rely on the existence of some preexisting underlying method to synchronize the data collected by devices used in fault localization. The quality and accuracy of all fault localization methods is, as a result, determined by an external factor, which is the synchronization method and underlaying infrastructure. The research presented in this dissertation presents a new method that not only utilizes multiple devices to resolve challenges unique to the distribution grid but also provides a unique synchronization technique. This synchronization technique (presented in Chapter 3) removes dependency on a pre-existing external synchronization scheme. The synchronization technique uses the same data collected by the devices for fault localization to synchronize the data across all devices. The fault localization method presented in Chapter 4 is built alongside this synchronization technique. This method represents a new development in the application of multiple devices for fault localization in the distribution grid. These two components create new a system that allows highly accurate fault localization using a new multi-device method which is independent of third-party data synchronization equipment and methods. The final contribution is the optimization of the fault localization method which is discussed in Chapter 5. This optimization is in regard to the locations and number of devices used for fault localization. The optimization technique leverages metrics established in the fault localization method to clearly determine the performance of any given set of device locations and determine the most effective locations for the devices.

1.3.3 Summary of Contributions

The following list highlights the contributions of each of the three elements that are contained in this dissertation:

1. Synchronization of devices – Here a novel method is provided that allows the synchronization of fault data across all devices used in a distribution grid once a switch or recloser is present. Existing work did not consider the synchronization of the data which is key to the accuracy of all fault localization methods. With this synchronization method all fault localization methods using travelling waves can now ensure that all data recorded is synchronized.

2. Fault localization using segmentation – in this element a fault localization method is presented which prevents the known issue of multiple fault locations being generated when localizing faults on a distribution grid. The work presented is able to guarantee a single fault location once there is sufficient sensor coverage of the distribution grid.

3. Optimization of sensor placement – here an optimization algorithm is provided that determines the best location for sensors used for the fault localization method developed. The work presented here lays the groundwork for optimizing sensor placement by creating a framework for efficiently evaluating solution performance.

1.4 Outline and Overview

The remainder of this thesis is divided into five chapters. Chapter 2 provides an extensive review of the research conducted in the focus areas of fault localization and synchronization techniques. Chapters 3 to 5 encompasses articles that address the three major contributions of this thesis. Chapter 6 provides a conclusion for all work presented and discusses the future work to be done. The articles in chapters 3 and 4 contain published articles written by the author during the course of conducting the research and developing the work presented in this dissertation. A statement of co-authorship for these articles is provided at the beginning of this thesis. An overview of each chapter is as follows.

Chapter 2: Literature Review

This chapter provides an extensive review of work done in the area of fault localization and device synchronization. Firstly, it builds on the background information introduced in Chapter 1.2 by discussing the general fault localization process. This chapter then looks at the initial solutions created for fault localization in the transmission grid. At this point the general categories which cover fault localization techniques are highlighted. After establishing the initial work and the categories covering fault localization specific focus is then placed on work for fault localization in distribution grids. The literature review then looks at work done in synchronization methods in Internet of Things (IoT). Finally, some limitations that should be considered in fault localization method development is discussed.

Chapter 3: Device Synchronization for Fault Localization in Electrical Distribution Grids

Chapter 3 provides an efficient method of synchronizing devices deployed in an electrical grid. The proposed method focuses on device synchronization specifically for localizing faults on distribution networks. It analyses the travelling waves that are present on the electrical grid at and around the time of the fault event. There are two sources of travelling waves that occur when a fault occurs. This chapter presents a synchronization method that synchronizes signals recorded by the sensors/devices and a recloser which emits a signal when a fault event has been detected. There is no reliance on accuracy of clocks used in each device or a pre-existing or third-party synchronization scheme. The proposed synchronization method resulted in a set of equations that characterize the delays between a fault signal and the recloser signal received at a sensor.

Chapter 4: Fault Localization in Smart Grids Using Segmentation

Chapter 4 describes a highly efficient method of localizing faults on electrical distribution grid while maintaining a high level of accuracy presented in [Hunte et al. 2021]. First the grid is broken down into segments, each with a unique identifier. The solution then analyses the travelling waves that are present on the electrical grid at the time of the fault. These signals are measured by devices at multiple locations and used to generate a key that is then utilized in conjunction with a segmented graphical model of the electrical distribution grid to determine the location of the fault.

Chapter 5: Optimization of Sensor Deployment in Electrical Distribution Grids for Fault Localization

Chapter 5 builds on the original work of localizing faults in distribution grids using graph segmentation as described in Chapter 4. The motivating goal is to optimize the placement and number of sensor devices used in the fault localization system developed in [Hunte et al. 2021]. It uses the new concept of primary and secondary localization segments established in Chapter 4 to evaluate the performance of any given solution. Two contrasting approaches to finding the best solution (device locations) are discussed and compared to determine an effective method of optimizing the placement of sensors in a problem that does not scale well. The first method is a greedy algorithm that restricts the possible solutions evaluated and significantly reduces the computational cost, and the second is a genetic algorithm that eases those restrictions at an increased computational cost.

1.5 Bibliography

Arnold, G. W., D. Wollman, G. FitzPatrick, D. Prochaska, D. G. Holmberg, D. H. Su, A. R. Hefner Jr, N. Golmie, T. L. Brewer and M. Bello. "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0." (2010).

Gabr, M. A., D. K. Ibrahim, E. S. Ahmed and M. I. Gilany. "A new impedance-based fault location scheme for overhead unbalanced radial distribution networks." *Electric Power Systems Research* **142**: 153-162, (2017).

Girgis, A. A., C. M. Fallon and D. L. Lubkeman. "A fault location technique for rural distribution feeders." *IEEE Transactions on Industry Applications* **29**(6): 1170-1175, (1993).

Gopakumar, P., M. J. B. Reddy and D. K. Mohanta. "Fault Detection and Localization Methodology for Self-healing in Smart Power Grids Incorporating Phasor Measurement Units." *Electric Power Components and Systems* **43**(6): 695-710, (2015).

Gopstein, A., C. Nguyen, C. O'Fallon, N. Hastings and D. Wollman. "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 4.0." *NIST Spec. Publ*, (2021).

Hoffert, M. I., K. Caldeira, G. Benford, D. R. Criswell, C. Green, H. Herzog, A. K. Jain, H. S. Kheshgi, K. S. Lackner, J. S. Lewis, H. D. Lightfoot, W. Manheimer, J. C. Mankins, M. E. Mauel, L. J. Perkins, M. E. Schlesinger, T. Volk and T. M. L. Wigley. "Advanced Technology Paths to Global Climate Stability: Energy for a Greenhouse Planet." *Science* **298**(5595): 981-987, (2002).

Hunte, J., H. Lutfiyya and A. Haque. Fault Localization in Smart Grids Using Segmentation. *ICC 2021 - IEEE International Conference on Communications*, (2021).

Industry-Partner. Ontario Electricity Provider, (2018).

Jun, Z., D. L. Lubkeman and A. A. Girgis. "Automated fault location and diagnosis on electric power distribution feeders." *IEEE Transactions on Power Delivery* **12**(2): 801-809, (1997).

LaCommare, K. H. and J. H. Eto. Understanding the cost of power interruptions to U.S. electricity consumers(2004).

Lazaropoulos, A. G. "Main Line Fault Localization Methodology (MLFLM) in Smart Grid – The Underground Medium- and Low-Voltage Broadband over Power Lines Networks Case." *Trends in Renewable Energy; Vol 4, No 1 (2018)*, (2017).

Magnago, F. H. and A. Abur. "Fault location using wavelets." *IEEE Transactions on Power Delivery* **13**(4): 1475-1480, (1998).

Myeon-Song, C., L. Seung-Jae, L. Duck-Su and J. Bo-Gun. "A new fault location algorithm using direct circuit analysis for distribution systems." *IEEE Transactions on Power Delivery* **19**(1): 35-41, (2004).

Pabla, A. S. Electric Power Distribution, McGraw-Hill Education, (2005).

Pourahmadi-Nakhli, M. and A. A. Safavi. "Path Characteristic Frequency-Based Fault Locating in Radial Distribution Systems Using Wavelets and Neural Networks." *IEEE Transactions on Power Delivery* **26**(2): 772-781, (2011).

Prajapat, D. "Faults and Effects in Electrical Power System." (2018). from https://madhavuniversity.edu.in/faults-and-effects-in-electrical-power-system.html.

Sapountzoglou, N., B. Raison and N. Silva. Fault Detection and Localization in LV Smart Grids. *2019 IEEE Milan PowerTech*, (2019).

Srinivasan, D., R. L. Cheu, Y. P. Poh, A. Kim and C. Ng. "Automated fault detection in power distribution networks using a hybrid fuzzy±genetic algorithm approach." (2000).

Takagi, T., Y. Yamakoshi, M. Yamaura, R. Kondow and T. Matsushima. "Development of a New Type Fault Locator Using the One-Terminal Voltage and Current Data." *IEEE Transactions on Power Apparatus and Systems* **PAS-101**(8): 2892-2898, (1982).

Wen, F. and C. S. Chang. "A new approach to fault diagnosis in electrical distribution networks using a genetic algorithm." *Artificial Intelligence in Engineering* **12**: 69-80, (1998).

Chapter 2

2 Background and Related Work

Presented in this chapter is a review of work done in the area of fault localization and device synchronization. This chapter first discusses the general fault localization process. Next a review of the initial solutions created for fault localization in the transmission grid is presented highlighting the general categories which cover the existing fault localization techniques. After establishing the initial work and the categories covering fault localization specific focus is then placed on work for fault localization in distribution grids. Next a review of synchronization methods in Internet of Things (IoT) is discussed. Finally, some limitations that should be considered in fault localization method development are highlighted.

2.1 Fault Localization

One of the initial thoughts on the fault localization problem was on a possible link to a computer network and thus the possibility that a fault localization solution could be developed based on a computer network. However, it may not be as promising an avenue of research as initially thought. This is due largely to a main fact that differentiates computer networks from the electrical grid. When a fault occurs on an electrical grid the power is disconnected from a large section. In a network if a fault occurs and a node goes down the network remains running and alternate pathways can be used to send information that can localize the bad node or area. This is not the case in an electrical grid. In an electrical grid power is turned off so no communication channels exist, and no nodes remain in operation to report information.

In order for a fault to be localized and fixed the presence of a fault must first be detected. With respect to fault detection there are two categories of faults. The scenarios are High Impedance Faults (HIF) and Low Impedance Faults (LIF). Traditionally the electrical grid used devices that would detect abnormal currents or voltages present on the line to detect faults. This technique is successful in detecting faults with low impedance. This is because these faults result in a high current that can be detected by a range of devices. With LIF a threshold can be set for the current and once the current passes the threshold then the fault is detected. These faults also cause the phases to become unbalanced which provides another well-used method of detecting faults (e.g. [Balser et al. 1986]) . It is highlighted by [Jota and Jota 1999] that HIFs are more difficult to detect. They are difficult to detect because a high impedance fault would mimic the normal operation of the grid. It is also noted that HIFs do not create a noticeable change in current and are therefore very difficult to detect [Jota and Jota 1999]. The paper presented by [Jota and Jota 1999] also states the only way these types of faults are detected traditionally is by customer call-ins due to power loss or sighting of down power lines. In relation to the quantity of HIFs [Wester 1998] states that only 5 – 20 percent of all faults are HIFs. The primary focus of research and development in the field has therefore been to localize LIFs which are responsible for up to 80 – 95 percent of faults. These faults are detected most often by using voltmeters and ammeters that record the voltage and current on the electrical grid. The current and voltage is altered by the fault and changes the expected 60Hz sign wave. The effects of faults and how they are measured is also discussed Section 2.1, 2.1 and Chapter 3.

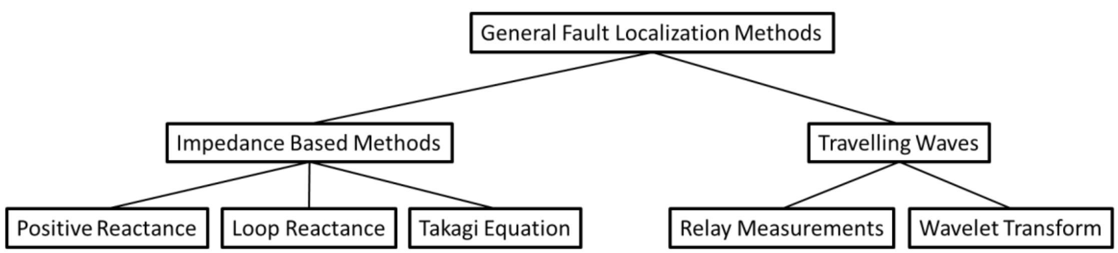

Figure 2.1. General fault localization techniques

Figure 2.1 shows the general fault localization methods as highlighted by [Marguet 2015]. Initially these methods were developed for the transmission grid as that was the initial focus of fault localization. These methods are still considered general because they are the primary methods used to localize faults that have been adapted for both the transmission and distribution grid with varying levels of success.

2.1.1 Fault Localization with Impedance based Equations

Within the impedance-based methods [Marguet 2015] identified three main sub-categories. These are positive reactance, loop reactance and the Takagi equation. These methods use equations that require precise details about the elements on the grid. These techniques require information that includes the type and the length of the conductor, and the amount of power being used by each consumer.

2.1.1.1 Positive Reactance

Positive reactance computation is used to estimate the fault distance in works by [Saha et al. 2001] and [Das et al. 2011]. A challenge specific to this approach is that it has to make an assumption on the load in the network. This can be particularly challenging as the amount of power being used by certain areas of the network is difficult to determine at a given time since customers' power consumption at any time is very dynamic. The powering on or off of components in the home effects the amount of power travelling through the network. Given that multiple customers exist in any section of the network, knowing the exact power consumption is usually very difficult. This leaves the accuracy of the fault location heavily dependent on estimations which is difficult to do because of the dynamicity of customer power consumption.

2.1.1.2 Loop Reactance

Loop reactance uses the loop impedance calculation to determine fault distance. This approach is presented by [Karnik et al. 2011]. This approach does not consider the current generated by the load in its calculation of the distance.

2.1.1.3 Takagi Equation

Takagi equation presented by [Takagi et al. 1982] uses the voltage and current measurements to compute the fault distance. This approach also requires that an assumption be made on the load on the network. The assumption in this case is the load in the network only occurs beyond the fault location and that the load also remains the same as its pre-fault value. This is inaccurate as there are many branches in the network and faults can occur anywhere. This means that in most cases there will also be a load between the fault and the measurement device. This assumption therefore almost always guarantees that the system is not represented accurately.

2.1.2 Fault Localization with Travelling Waves

When a fault occurs in the grid a wave is emitted from the fault location and travels across the grid like a ripple in a pond. This *incident* wave reflects off the end of the line and travels back up the line. Figure 2.2 shows this phenomenon on a single line network.

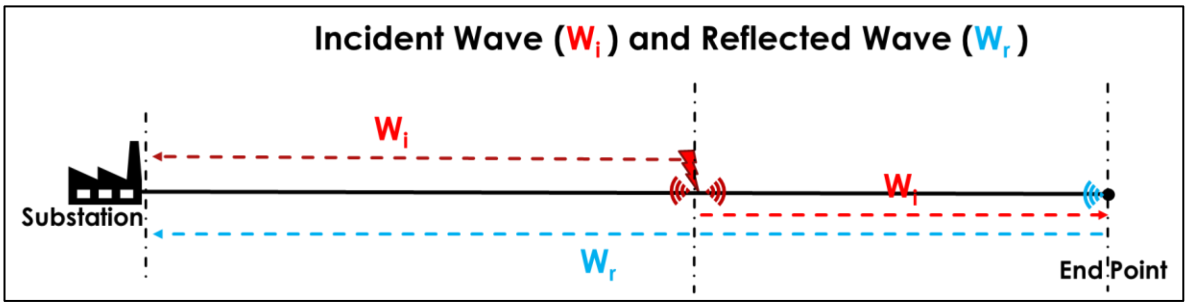

Figure 2.2. Incident and reflected waves in an electrical grid

When a fault occurs in the grid a wave is emitted from the fault location and travels across the grid. This *incident* wave reflects off the end of the line and then travels back up the line. The reflected waves also bounce off the fault location. Figure 2.3 shows this occurring on a single line network. The sensor S_1 records the arrival times of waves on the line between S_1 and the endpoint (EP_1). This list of arrival times makes up the fault signal. When the fault occurs at time T_0 the incident wave W_i travels up the line and arrives at the sensor S_1 at T_1. The incident wave also travels down the line and reflects off the endpoint and creates the reflected wave W_r. This reflected wave W_r travels back up the line and arrives at the sensor at time T_2. There are a number of other reflections that bounce off different combinations of the fault and the endpoint.

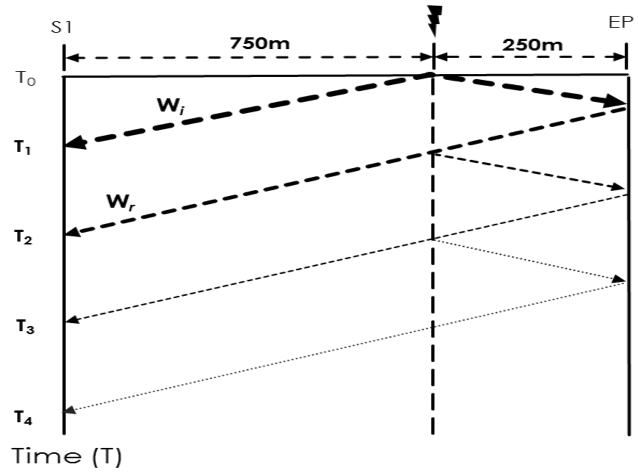

Figure 2.3. Technical diagram of travelling waves

The work presented in [Marguet 2015] sub-categorizes travelling wave methods into relay measurement and wavelet transform methods.

2.1.2.1 Relay Measurements

Relay measurements are used to detect fault surges and calculate distances using a given propagation delay. An example of the use of relay measurements can be found in [Crossley and McLaren 1983]. The work done in [Crossley and McLaren 1983] represents one of the first uses of measured currents and voltages to locate faults. Relay measurements record the current and voltage data on the network. This data also contains the fault surges that were discussed previously. Once the time of these surges is also captured the distance from the fault can be calculated once the speed that these surges travel is known. According to [Crossley and McLaren 1983] the relays must be able to identify the incident surge and the reflection of the incident wave. Once these can be identified the times they are detected can be logged. Part of the work presented by [Crossley and McLaren 1983] explains that by taking the delay between these two waves and multiplying it by the propagation velocity of the waves, the distance from the relay point to the fault can be determined. Relay measurements then lead to the development of wavelet transform methods. This was done to address the complications of detecting the surges within the existing waveforms present on the electrical grid. Other work ranges from studying different techniques that use relays and wavelets as seen in [Bumanapalli Ravindranath et al. 2007] to studying the effectiveness of methods depending on the type of electrical networks seen in [Achleitner et al. 2008].

2.1.2.2 Wavelet Transform

Wavelet transform methods build on the original relay measurement technique. They improve relay measurements by using a wavelet transformation on the signal to extract high frequency surges and then calculate distances using a given propagation speed. An example of a wavelet transform method is presented by [Magnago and Abur 1998]. The challenge in using this method as highlighted by [Marguet 2015] is that it requires a high sampling rate. This travelling wave approach is often used in the transmission grid. A fault localization technique is presented in [Hizam and Crossley 2006] that uses a model of the network and a simulation of faults to perform the localization. The solution proposed by [Hizam and Crossley 2006] uses the known reflection points of the network and the time delays between travelling waves traversing the network when a fault occurs. The final location of the fault requires a simulation of possible locations in a simulator and a subsequent comparison for similarity is required for the final localization.

2.2 Fault Localization in Distribution Grids

Transmission grids are effectively the intercity backbone responsible for delivering power to cites. If a fault occurs in the transmission grid it would therefore result in power loss to entire cities. Hence, initial fault localization techniques focused on these networks.

After successfully addressing the localization of faults in the transmission grid focus then shifted on further improving the continuity of power service within the smaller impact area- the distribution grids. Initially attempts were made to apply transmission grid solutions within the distribution grid but this had limited results.

The main issues found in these techniques are with the increased complexity of the distribution grid. These characteristics make fault localization solutions developed for transmission grids difficult to migrate to the distribution grid. Table 2.1 shows the characteristics and differences between the transmission and distribution grid.

Table 2.1 Characteristics of the transmission and distribution grids.

	Characteristics	**Transmission Grid**	**Distribution Grid**
1	Area of coverage	~ 1000s of miles	~100s of miles
2	Number of branches	Very low (< 10)	Very high
3	Number of components	Few	Many
4	Changes made to structure	Very Few	Many (Dynamic)
5	Accuracy of data on structure	High	Estimated
6	Variations on conductors	None	Yes
7	Location (above/below) ground	Above	Above and below
8	Structure of grid	Straight Line (End to End) and Radial	Straight Line, Radial and Looped

The following list provides an explanation of the 8 characteristic differences outlined in Table 2.1.

1. The area covered by the distribution grid is smaller making small errors in localization much more complicated to correct.
2. Distribution grids have many branches on any given feeder.
3. There is a substantially a higher number of components present in distribution grids than in transmission grids, complicating calculations made for localization.
4. The structure of the distribution grid is much more dynamic as new customers are added much more often on the distribution grid, changing its structure.
5. The information available for the electrical grid typically has more errors as its components are changed and new ones added.
6. The distribution grid can have any number of different conductor types as several transformations are made on the voltage to step it down to consumer usage needs.
7. Distribution networks have potions of their network running underground. This changes the properties of the variables used in some fault localization calculations. For

example, the impedance due to a change in conductor type, as these cables are insulated and cannot carry the same amount of power due to lack of heat dissipation.

8. The transmission grid largely focuses on one-way transfer of power from upstream (power generation plants) to downstream (cities). It is mostly an end to end transfer of power with radial branches along this path that supply cities along the path. In the distribution grid, there is a much higher complexity in power flow as there are frequent loops in its structure. There are also switches that may cause the flow of power along a section of the network to be reversed or to take a different path entirely.

The distribution grid is therefore seen as a much more complex system due to these characteristics highlighted in Table 2.1. These characteristics make fault localization much more difficult. This is also discussed by [Marguet 2015]. There have been a number of implementations of the general solutions shown in figure 8 in the distribution grids. In the next section, a focus is placed on solutions designed specifically for distribution grids, in an attempt to handle the complexity of the distribution grid.

To address the challenges in applying the general methods to the distribution grid, a number of solutions were developed. These solutions include the use of:

- An enhanced analysis of travelling waves
- Fuzzy set theory
- Direct circuit analysis
- Iterative fault distance
- Impedance-based methods
- Phasor measurement units
- Smart meters
- Current Sensors

This is by no means an exhaustive list of the methods developed but only seeks to capture a variety of these methods. The methods in this list represent a sample of some of the more common approaches taken to localize faults in distribution grids.

2.2.1 Enhanced Analysis of Travelling Waves

In the work associated with fault localization in distribution grids, a particular focus has been placed on enhancing the solutions that utilize travelling waves to allow them to be more accurate on the distribution grid. These solutions look to add more sophisticated analysis of the travelling waves and add new components to improve accuracy of the localization results. These techniques employ neural networks and genetic algorithms. Fault cases of fault locations and corresponding energy vectors are created in [Pourahmadi-Nakhli and Safavi 2011] that are unique to the path to the fault on a given network. These fault locations and energy vector pairs are used to train a neural network to select the fault location based on a given energy vector. The works of [Wen and Chang 1998, Srinivasan et al. 2000] use a genetic algorithm as part of a solution that provides support for handling more complicated faults and locate faulty sections and defects in the feeder protection systems at the same time. Though the approaches show an increased performance, the dynamic nature of the distribution grid may pose a major challenge to accessing and maintaining a record of the data that is needed to establish the energy vectors.

2.2.2 Fuzzy Set Theory

An approach that uses weather is investigated by [Jarventausta et al. 1994]. They look to weather as a strong influencer on faults. Their solution uses the knowledge of fault types according to the weather and the physical locations in which sections of the electrical grid are located. According to [Jarventausta et al. 1994] fuzzy set theory allows decisions to be made in the absence of statistical data. Instead of determining the probability of a fault location it determines the possibility of a fault being in that location. It uses the geology of the areas around the network to determine if a location could have a fault. The use of data on weather and fuzzy set theory to localize faults can be effective. However, this requires access to and maintenance of data like geological records of the area surrounding the distribution grid which may not always be available or accurate.

2.2.3 Direct Circuit Analysis

The work of [Myeon-Song et al. 2004] focuses on improving an equation-based solution. The traditional solution uses direct circuit analysis. It is also stated by [Myeon-Song et al.

2004] that an unbalanced grid provides complications that prevent traditional direct circuit analysis equations from being used. It is then argued by [Myeon-Song et al. 2004] that the derivation of the fault location equation becomes too complex in an unbalanced system. Their solution introduces a method that simplifies the derivation process for the conventional fault localization equation. The simplification is achieved by applying a matrix inverse lemma. The direct circuit analysis approach's major challenge is in the complexity of the calculations used to determine the fault location. Even though the complications were addressed, the data needed creates a potential point of failure if the accuracy of the data is not precise in an already complex process.

2.2.4 Iterative Fault Distance

This solution developed by [Jun et al. 1997] uses steady state analysis of the distribution network. It uses known characteristics about the distribution feeder (the cables used in a distribution grid to transport power) and information recorded by electrical voltage and current measurement devices when the fault occurs. This method uses a model of the system and would therefore have potential inaccuracies. To address the potential inaccuracies of the model, the algorithm estimates fault regions using probabilistic modeling and analysis. The method first computes possible fault locations using readings at the substation. These possible locations are then ranked and pruned to select the actual fault location using a fault diagnosis algorithm. This method can be effective at addressing the complications of the distribution grid. However, it would require constructing a complex probabilistic model and a fault diagnosis algorithm. An accurate model would require ongoing maintenance as adjustments would need to be made when changes are made to any components which would become very challenging if the grid is providing power in a developing area.

2.2.5 Impedance-Based Methods for Distribution Grids

The use of voltage and current measurements for fault localization in the distribution grid was initially examined by [Girgis et al. 1993], [Jun et al. 1997] and [Seung-Jae et al. 2004]. The work in [Gabr et al. 2017] adds to this work on applying impedance based methods to the distribution grid. [Gabr et al. 2017] uses single ended voltage and current measurements and uses a quadratic formula to estimate the location of all types of faults. The work in [Gazzana et al. 2014] adds to this work by developing a hybrid solution made up of impedance and travelling wave approaches where measurements are single ended (at the substation). However [Mwifunyi et al. 2019] notes that these methods use constant impedance load modelling which is somewhat detached from the variability of the load in a real distribution grid. The paper presented by [Mwifunyi et al. 2019] also highlights that these impedance-based methods main shortfall is that they provide multiple estimations of possible fault locations. An attempt is made by [Estebsari et al. 2016] to address the main shortfall of multiple estimations of locations by developing another hybrid technique of impedance-based and voltage sag methods to reduce the number of estimations.

The use of Power-line Communication modems (PLC modems) is introduced in [Passerini and Tonello 2017]. Power-line communication is used to send data over the same cables that distribute power throughout an electrical grid. the work in [Passerini and Tonello 2017] seeks to leverage the presence of PLC modems in grids that utilize PLC. As the PLC modems will already be deployed in such a distribution grid [Passerini and Tonello 2017] look at the employment of these devices to detect high impedance faults (HIF). Focus is placed on HIFs by [Passerini and Tonello 2017] as they are more difficult to detect (as discussed in section 4) but can still be a public hazard. The solution uses a single impedance measurement at the central office. The work done by [Lazaropoulos 2017], [Lazaropoulos 2017] and [Lazaropoulos 2017] also looks at the development of a fault localization technique using a PLC method known as broadband over power lines (BPL). These are mentioned as they are a fault localization technique but are only possible in a power grid that uses PLC equipment which is very uncommon.

2.2.6 Phasor Measurement Units

Phasor Measurement Units (PMUs) are employed by [Gopakumar et al. 2015] throughout the grid to access voltage and current data. Their work primarily attempts to localize faults in the transmission grid and then briefly looks at applying the PMUs in the distribution grid. The solution created by [Gopakumar et al. 2015] analyses the voltage phasor angle at the bus using a fast Fourier transform. A bus in this context refers to the junction on a distribution feeder where the distribution feeder cable branches off to supply multiple areas with power. A three step localization technique is introduced by [Gopakumar et al. 2015]. The first is the identification of the bus related to the faulty branch. This is done by looking at the phase angle deviations of the busses. The second is the identification of the faulty branch connected to that bus, done by a comparison of all the busses connected to that faulty bus. The final step is the distance from the bus to the fault along the identified faulty branch. This final step is done by analyzing the variations in the current phasor angle in the faulty branch.

The work done by [Jamei et al. 2018] and [Jamei et al. 2020] focus on determining the optimal placement of PMUs. In [Jamei et al. 2018] an initial look at the limitations of fault localization using PMUs is discussed. The work presented in [Jamei et al. 2018] also introduces the notion of a fault location being calculated as a group of locations where determining a specific fault location may not be possible. The effect of the placement of sensors on fault localization is determined by [Jamei et al. 2020] who offer some initial insight on the topic and continue their work on PMU placement. The work in [Jamei et al. 2020] also analyzes the distribution of different fault location hypotheses using the Kullback Leibler divergence and analyze the topology of the grid and how the placement of PMUs effect the ability to localize faults. The goal is to use this analysis to formulate a PMU placement strategy for a given number of sensors. The use of multiple devices and the optimization of their placement in the grid shows a very promising direction for future work. However, there may be some difficulty with the use of the variations in phasor angle as the phasor angle can change quite frequently.

2.2.7 Smart Feeder Meters and Voltage Sag

The use of meters for the distribution feeder is looked at in [Trindade et al. 2014] and [Trindade and Freitas 2017]. The introduction of meters for feeders that are able to monitor voltage sag is made by [Trindade et al. 2014]. In their work, [Trindade et al. 2014] combine the voltage sag data with outage maps to localize faults on distribution feeders. They then build on this in [Trindade and Freitas 2017] to enhance the accuracy of their original idea by determining the best location for the feeder meters.

The presentation of a new approach to localize faults using a three-dimensional mapping of phase voltage as a polarization ellipse is made by [Alam et al. 2019]. Within [Alam et al. 2019] they explain that fault classification is carried out by the decision boundaries set by five parametric equations. Localization of faults is done using the five parametric equations by using the variations in these five parameters with respect to distance from the fault location.

A solution is discussed by [Škumát and Ž 2019] specifically for open circuit fault localization utilizing smart meters. The solution presented by [Škumát and Ž 2019] uses the smart meters to capture voltage data in order to localize open circuit faults. The open circuit fault causes a change in voltage that the smart meters can detect if that meter is beyond the fault. The fault location is said to be between the last meter detecting a normal voltage and the first meter detecting an open circuit fault. The work in [Škumát and Ž 2019] notes that the use of smart meters in the fault localization process is a very promising concept for medium and low voltage grids. This is an efficient approach as it does not require modelling or external data sources. However, it depends on the distance between the smart meters which limits its accuracy when pinpointing a fault's location.

2.2.8 Current Sensors and Voltage Drop

A fault localization method that uses voltage measurements is presented in [Sapountzoglou et al. 2019]. The localization of a fault is performed as a three-step process of faulty branch identification, faulty sector localization and a distance estimation from the beginning of the feeder. The voltage data used for this process differs depending on if the fault is a single-phase to ground or a three-phase fault. Phase voltage was used for single-phase to ground faults and positive sequence component of the voltage was used for three-phase fault cases. Six factors are considered by [Sapountzoglou et al. 2019] when developing the solution. These include the unbalanced nature of the grid; faults at different points in a feeder (beginning, middle or end); different fault types; different load situations; different fault resistances and the placement of sensors. [Sapountzoglou et al. 2019] show that the use of sensors can provide reliable solutions for fault localization in distribution grids. This is a very effective approach and considers many of the challenges faced in distribution grid fault localization. The limitation of this solution is in the use of voltage measurements which are difficult to capture in an electrical grid.

2.3 Synchronization in Internet of Things

The solutions developed for fault localization in the distribution grid has gradually moved towards the use of multiple devices throughout the grid. As such, a review of synchronization methods was deemed necessary in order to utilize the data from multiple sensors effectively. The fault events in electrical distribution grids occur in fractions of seconds. The signals used in the solutions propagate through the distribution grid at speeds approaching the speed of light. Many of the approaches mentioned rely on the recorded timing of these fault events. It is therefore imperative that the timing of these events be as accurate as possible. When using more than one device, this would require that the clocks of these devices be synchronized so that an accurate log of fault events can be compiled. Clock synchronization can be achieved using a number of methods:

- **Berkeley algorithm** – This technique developed by [Gusella and Zatti 1989] uses the average time between multiple nodes as the correct time and all nodes are then adjusted to reflect this time.
- **Clock-sampling mutual network synchronization** – This technique developed by [Rentel and Kunz 2005] uses the fact that every clock in each node has a time drift factor. Once this factor is calculated by a node then their clocks can be corrected.
- **Cristian's algorithm -** This method was developed by [Cristian 1989] and relies on the existence of an accurate time source and a time server. Clients then use the time server to retrieve the correct time.
- **Global Positioning System -** GPS satellites have atomic clocks that allow them to be extremely accurate. The GPS satellite sends a signal which allows a receiving device to by synchronized to UTC.
- **Network Time Protocol** – A very widely used method for achieving millisecond accuracy in unreliable networks and is used across the internet.
- **Precision Time Protocol** – A master slave method of delivering highly accurate time that is used mainly for synchronization in local area networks.
- **Synchronous Ethernet** – This technique transmits synchronization signals over the ethernet physical layer.

With each the above methods except the GPS method, a preexisting network is required through which the synchronization protocols are executed. Our focus is on an IoT system where in our case, there may not be such a network available and where communication throughout the system is expensive and not always available. As such, a specific look was taken at synchronization in IoT. According to [Viswanathan et al. 2016] there are two main categories of synchronization specifically for IoT applications.

1. Using the exchange of messages between the nodes of the system.
2. Using periodic external signals (that are not part of the system).

Unfortunately, the first IoT synchronization category also relies on a preexisting underlaying network. We still present the techniques in this category for completeness in section 2.3.1. The external signal methods presented in section 2.3.2 are of particular interest to our work given the lack of a network or high cost in communication.

2.3.1 Internal Message Synchronization

Internal messages synchronization methods are techniques that allow nodes on the network to send messages to each other that enable them to synchronize their clocks. The Reference-Broadcast Synchronization (RBS) scheme is presented in [Elson et al. 2002]. The message sent in this scheme does not provide a timestamp. Instead, the arrival time of the messages are used to synchronize clocks. A two-step protocol referred to as Timing-sync Protocol for Sensor Networks (TPSN) is presented in [Ganeriwal et al. 2003]. The initial step determines a hierarchy and then pairwise synchronization is performed among the nodes, ensuring that all nodes are synchronized to the reference node at the top of the hierarchy. During the presentation of the work in [Ganeriwal et al. 2003] they argue that this method performs twice as good as RBS. The Flooding time Synchronization Protocol (FTSP) is studied in [Maroti et al. 2004]. This protocol is resistant to node failures and achieves this by dynamically updating the topology of the node network and periodically flooding the system with synchronization messages. The discussion in [Maroti et al. 2004] explains that their protocol's performance is achieved by utilizing MAC-layer time stamping, and error compensation using clock skew estimation. They state that their solution outperforms both the RBS and TPSN methods by achieving and average per-hop error of around one microsecond. The solutions in this category require regular communication with all nodes.

These solutions also require an existing network between all nodes. Unfortunately, the network used for this method is not available in most electrical grids.

2.3.2 External Signal Synchronization

There are a number of external signals that have been used as time keeping events in order to allow devices that can detect them to synchronize their clocks. Time keeping radio stations signals are used by [Chen et al. 2011]; [Li et al. 2011] use the Radio Data System; [Hao et al. 2011] take advantage of Wi-Fi beacons using Zigbee nodes to detect the beacons; [Rowe et al. 2009] design a device that utilizes electromagnetic radiation to calibrate clocks by detecting ac cycles; [Li et al. 2012] employs light sensors to detect the flickering of fluorescent lights and calibrate clocks. The solutions that utilize external signals allow devices to operate autonomously. These solutions do not explicitly require an existing network as the synchronization signal is not passed through the network. These methods take advantage of signals that already exist in their environment, which allows the system to synchronize without network communication.

2.4 Limitations of Existing Work

An initial review of the general fault localization problem and existing solutions was chosen as the starting point. This provides the path that the research has taken so far and provides information on where it is going. [Marguet 2015] discussed the initial motivation of fault localization and its origins in the transmission grid and looked at its migration over time into the distribution grid. The most important factor discovered upon such a review was the foundation provided by initial solutions for fault localization. These fundamental approaches can be summarized into two main types:

1. Temporal analysis of transient signals created by the fault event.
2. Comparative analysis of measurements of power variables (current, voltage, load etc.) during the fault event.

It can be argued that all fault localization techniques fall under either of these categories or contains elements of one or both categories. These two fundamental categories to fault localization solutions would remain as the primary methods from which solutions emerge. In recent research, it was found that newer solutions applied various enhancements to improve performance and applicability. These enhancements have ranged from improvements in physical devices and their capabilities to enhanced methods of analyzing fault signals and data.

The transient signal category (travelling wave approaches) has the major challenge of handling the exceptionally complex signals present on the distribution grid. The transient signals in the distribution grid bounce off all the distribution grid's many components (like meters, switches, transformers line terminations etc.). This is highly challenging as there are hundreds of these components on a single feeder in the distribution grid and these reflected signals can cancel out each other.

The power variable category also suffers due to the complexity of the distribution grid. Many of these variables are highly dynamic and are impacted by every consumer drawing power from the grid. There is also the presence of noise and effects of environmental influencers (temperature, humidity, animals, tree branches etc.). Even though these may have a smaller impact they can reduce the accuracy of the calculations.

Another strong limitation of fault localization in electrical grids is in testing. To implement many of these solutions in the grid for testing, new equipment must be added to electrical grid. This poses quite a challenge as there is risk involved in adding new components to the electrical grid that may adversely affect its consumers. In addition to this, many technical problems would have to be solved like determining power sources for the devices and communication methods between the devices. As a result of this, few solutions are placed into the electrical grid and tested in the field. Instead, simulations of the electrical grid are run to mitigate the risk involved by field testing. With these limitations in mind the solution should also draw the least risk and be cost effective when being implemented in an electrical grid.

The fault localization methods that have been developed for the distribution grid have had varying degrees of success. Some solutions have been quite successful in enhancing the localization methods used in the transmission grid like the impedance-based and travelling waves methods discussed in this chapter. These solutions have generally not been deployed due to the complex data requirements that may not be available. The more recent work in fault localization shows a movement toward the use of multiple sensors for fault localization. However, the challenge becomes keeping these devices synchronized as clock drift has a significant impact on the accuracy due to the speed of travel of signals on the grid. In conclusion, these multi device solutions are not perfect but indicate that a promising direction for future work is toward the use of multiple meters or sensors throughout the grid.

2.5 Summary

This chapter (Chapter 2) discusses fault localization methods in both the transmission and distribution grids. All fault localization approaches generally fall under two categories, Travelling Wave methods and Impendence-Based methods. This chapter started by examining the older approaches to localize faults in the transmission grid. These original approaches lay the foundation for all the work that follows for fault localization in distribution grids. The enhancement of the original travelling wave methods are the most successful methods developed for distribution grids. However, this success depends on the use of data captured using multiple devices. It was noted however that there is no consideration for the synchronization of these devices. This dissertation notes the synchronization of data across devices as a key element for the development of a fault localization solution. The next chapter (Chapter 3) is dedicated to the process of synchronizing the data used for fault localization. Chapter 3 presents a method to ensure the data collected for fault localization with travelling wave methods is accurately synchronized across all devices.

2.6 Bibliography

Achleitner, G., C. Obkircher, L. Fickert, M. Sakulin and C. Raunig. An earth fault distance location algorithm in compensated networks with additional estimation of the fault impedance and fault current. *2008 Power Quality and Supply Reliability Conference*, (2008).

Alam, M. R., K. M. Muttaqi and T. K. Saha. "Classification and Localization of Fault-initiated Voltage Sags using 3-D Polarization Ellipse Parameters." *IEEE Transactions on Power Delivery*: 1-1, (2019).

Balser, S. J., K. A. Clements and D. J. Lawrence. "A Microprocessor-Based Technique for Detection of High Impedance Faults." *IEEE Transactions on Power Delivery* **1**(3): 252-258, (1986).

Bumanapalli Ravindranath, R., C. P. Babu, M. Swetha, S. Munagala and B. P. Singh. Localization of faults in a transmission line using wavelet techniques. *2007 Annual Report - Conference on Electrical Insulation and Dielectric Phenomena*, (2007).

Chen, Y., Q. Wang, M. Chang and A. Terzis. Ultra-low power time synchronization using passive radio receivers. *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, (2011).

Cristian, F. "Probabilistic clock synchronization." *Distributed Computing* **3**(3): 146-158, (1989).

Crossley, P. A. and P. G. McLaren. "Distance Protection Based on Travelling Waves." *IEEE Transactions on Power Apparatus and Systems* **PAS-102**(9): 2971-2983, (1983).

Das, S., S. Kulkarni, N. Karnik and S. Santoso. Distribution fault location using short-circuit fault current profile approach. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Elson, J., L. Girod and D. Estrin. "Fine-grained network time synchronization using reference broadcasts." *SIGOPS Oper. Syst. Rev.* **36**(SI): 147-163, (2002).

Estebsari, A., E. Pons, E. Bompard, A. Bahmanyar and S. Jamali. An improved fault location method for distribution networks exploiting emerging LV smart meters. *2016 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*, (2016).

Gabr, M. A., D. K. Ibrahim, E. S. Ahmed and M. I. Gilany. "A new impedance-based fault location scheme for overhead unbalanced radial distribution networks." *Electric Power Systems Research* **142**: 153-162, (2017).

Ganeriwal, S., R. Kumar and M. B. Srivastava. Timing-sync protocol for sensor networks, ACM**:** 138-149, (2003).

Gazzana, D. S., G. D. Ferreira, A. S. Bretas, A. L. Bettiol, A. Carniato, L. F. N. Passos, A. H. Ferreira and J. E. M. Silva. "An integrated technique for fault location and section identification in distribution systems." *Electric Power Systems Research* **115**: 65-73, (2014).

Girgis, A. A., C. M. Fallon and D. L. Lubkeman. "A fault location technique for rural distribution feeders." *IEEE Transactions on Industry Applications* **29**(6): 1170-1175, (1993).

Gopakumar, P., M. J. B. Reddy and D. K. Mohanta. "Fault Detection and Localization Methodology for Self-healing in Smart Power Grids Incorporating Phasor Measurement Units." *Electric Power Components and Systems* **43**(6): 695-710, (2015).

Gusella, R. and S. Zatti. "The accuracy of the clock synchronization achieved by TEMPO in Berkeley UNIX 4.3BSD." *IEEE Transactions on Software Engineering* **15**(7): 847-853, (1989).

Hao, T., R. Zhou, G. Xing and M. Mutka. WizSync: Exploiting Wi-Fi Infrastructure for Clock Synchronization in Wireless Sensor Networks. *2011 IEEE 32nd Real-Time Systems Symposium*, (2011).

Hizam, H. and P. A. Crossley. Single Ended Fault Location Technique On a Radial Distribution Network Using Fault Generated Current Signals, (2006).

Jamei, M., R. Ramakrishna, T. Tesfay, R. Gentz, C. Roberts, A. Scaglione and S. Peisert. "Phasor Measurement Units Optimal Placement and Performance Limits for Fault Localization." *IEEE Journal on Selected Areas in Communications* **38**(1): 180-192, (2020).

Jamei, M., A. Scaglione and S. Peisert. Low-Resolution Fault Localization Using Phasor Measurement Units with Community Detection. *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, (2018).

Jarventausta, P., P. Verho and J. Partanen. "Using fuzzy sets to model the uncertainty in the fault location process of distribution networks." *IEEE Transactions on Power Delivery* **9**(2): 954-960, (1994).

Jota, P. R. S. and F. G. Jota. "Fuzzy detection of high impedance faults in radial distribution feeders." *Electric Power Systems Research* **49**: 169-174, (1999).

Jun, Z., D. L. Lubkeman and A. A. Girgis. "Automated fault location and diagnosis on electric power distribution feeders." *IEEE Transactions on Power Delivery* **12**(2): 801-809, (1997).

Karnik, N., S. Das, S. Kulkarni and S. Santoso. Effect of load current on fault location estimates of impedance-based methods. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Lazaropoulos, A. G. "Main Line Fault Localization Methodology (MLFLM) in Smart Grid – The Underground Medium- and Low-Voltage Broadband over Power Lines Networks Case." *Trends in Renewable Energy; Vol 4, No 1 (2018)*, (2017).

Lazaropoulos, A. G. "Main Line Fault Localization Methodology in Smart Grid – Part 1: Extended TM2 Method for the Overhead Medium-Voltage Broadband over Power Lines Networks Case." *Trends in Renewable Energy; Vol 3 (2017): SPECIAL ISSUE ON SMART GRID (2)*, (2017).

Lazaropoulos, A. G. "Main Line Fault Localization Methodology in Smart Grid – Part 3: Main Line Fault Localization Methodology (MLFLM)." *Trends in Renewable Energy; Vol 3 (2017): SPECIAL ISSUE ON SMART GRID (2)*, (2017).

Li, L., G. Xing, L. Sun, W. Huangfu, R. Zhou and H. Zhu. Exploiting FM radio data system for adaptive clock calibration in sensor networks, ACM**:** 169-182, (2011).

Li, Z., C. Li, W. Chen, J. Dai, M. Li, X.-Y. Li and Y. Liu. Clock calibration using fluorescent lighting, ACM**:** 463-466, (2012).

Magnago, F. H. and A. Abur. "Fault location using wavelets." *IEEE Transactions on Power Delivery* **13**(4): 1475-1480, (1998).

Marguet, R. "Improved fault localization method for electrical power distribution networks." (2015).

Maroti, M., B. Kusy, G. Simon and A. Ledeczi. The flooding time synchronization protocol, ACM**:** 39-49, (2004).

Mwifunyi, R. J., M. M. Kissaka and N. H. Mvungi. "Distributed approach in fault localisation and service restoration: State-of-the-Art and future direction." *Cogent Engineering* **6**(1): 1628424, (2019).

Myeon-Song, C., L. Seung-Jae, L. Duck-Su and J. Bo-Gun. "A new fault location algorithm using direct circuit analysis for distribution systems." *IEEE Transactions on Power Delivery* **19**(1): 35-41, (2004).

Passerini, F. and A. M. Tonello. Power line fault detection and localization using high frequency impedance measurement. *2017 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, (2017).

Pourahmadi-Nakhli, M. and A. A. Safavi. "Path Characteristic Frequency-Based Fault Locating in Radial Distribution Systems Using Wavelets and Neural Networks." *IEEE Transactions on Power Delivery* **26**(2): 772-781, (2011).

Rentel, C. H. and T. Kunz. A clock-sampling mutual network time-synchronization algorithm for wireless ad hoc networks. *IEEE Wireless Communications and Networking Conference, 2005*, (2005).

Rowe, A., V. Gupta and R. Rajkumar. Low-power clock synchronization using electromagnetic energy radiating from AC power lines, ACM**:** 211-224, (2009).

Saha, M., F. Provoost and E. Rosolowski. "Fault location method for MV cable network." (2001).

Sapountzoglou, N., B. Raison and N. Silva. Fault Detection and Localization in LV Smart Grids. *2019 IEEE Milan PowerTech*, (2019).

Seung-Jae, L., C. Myeon-Song, K. Sang-Hee, J. Bo-Gun, L. Duck-Su, A. Bok-Shin, Y. Nam-Seon, K. Ho-Yong and W. Sang-Bong. "An intelligent and efficient fault location and diagnosis scheme for radial distribution systems." *IEEE Transactions on Power Delivery* **19**(2): 524-532, (2004).

Škumát, T. and E. Ž. Open Phase Fault Localization in a Medium Voltage Distribution Grid Using Smart Meters. *2019 20th International Scientific Conference on Electric Power Engineering (EPE)*, (2019).

Srinivasan, D., R. L. Cheu, Y. P. Poh, A. Kim and C. Ng. "Automated fault detection in power distribution networks using a hybrid fuzzy±genetic algorithm approach." (2000).

Takagi, T., Y. Yamakoshi, M. Yamaura, R. Kondow and T. Matsushima. "Development of a New Type Fault Locator Using the One-Terminal Voltage and Current Data." *IEEE Transactions on Power Apparatus and Systems* **PAS-101**(8): 2892-2898, (1982).

Trindade, F. C. L. and W. Freitas. "Low Voltage Zones to Support Fault Location in Distribution Systems With Smart Meters." *IEEE Transactions on Smart Grid* **8**(6): 2765-2774, (2017).

Trindade, F. C. L., W. Freitas and J. C. M. Vieira. "Fault Location in Distribution Systems Based on Smart Feeder Meters." *IEEE Transactions on Power Delivery* **29**(1): 251-260, (2014).

Viswanathan, S., R. Tan and D. K. Y. Yau. Exploiting Power Grid for Accurate and Secure Clock Synchronization in Industrial IoT. *2016 IEEE Real-Time Systems Symposium (RTSS)*, (2016).

Wen, F. and C. S. Chang. "A new approach to fault diagnosis in electrical distribution networks using a genetic algorithm." *Artificial Intelligence in Engineering* **12**: 69-80, (1998).

Wester, C. G. High impedance fault detection on distribution systems. *1998 Rural Electric Power Conference Presented at 42nd Annual Conference*, (1998).

Chapter 3

3 Device Synchronization for Fault Localization in Electrical Distribution Grids

3.1 Introduction

This chapter focuses on the synchronization of data used in travelling wave fault localization methods. The most promising method of localizing faults in a distribution grid employs data from multiple sensors/devices in the localization process. However previous work does not consider the need for data synchronization across these devices. This chapter presents a method that synchronizes the travelling wave data recorded for fault localization It leverages a device known as a recloser that is part of the distribution grid.

The electrical grid consists of power plants, the *transmission network* and the *distribution network*. Power plants traditionally generate power from either coal, water, oil or nuclear. Power plants are typically located miles away from power consumers. In order to transmit the power over these large distances to consumers the power is sent over the transmission network through a transmission substation that converts the power to a high voltage suitable for long distance transmission. Once it reaches a cluster of electricity consumers it then uses the *distribution* network to deliver the power to the consumers. The transition from the transmission network to the distribution network is through a distribution substation which uses transformers to reduce the power voltage before it is transmitted to the consumers. A distribution feeder refers to the cables that transmit the power from substations.

Faults can occur at different stages throughout the electrical grid. These faults, given the amount of power passing through the grid, can cause massive damage to the equipment and even cause harm to people. Should a fault occur, the grid is designed to quickly stop sending power to that section of the grid. This protects people, the electrical grid, and its equipment.

To restore the power, the fault must first be localized. After this the fault can then be cleared and power can be restored. Faults can occur in both the transmission and distribution segments of the electrical grid. However due to the large area typically effected by a transmission grid level fault, initial localization techniques focused on the transmission grid. There are two categories of fault localization techniques [Marguet 2015]: impedance-based (e.g. [Saha et al. 2001],[Das et al. 2011], [Karnik et al. 2011], [Takagi et al. 1982]) and travelling waves (e.g. [Crossley and McLaren 1983], [Magnago and Abur 1998], [Hizam and Crossley 2006]).

For distribution grids, initial attempts focused on applying transmission grid solutions within the distribution grid. This had limited results since the distribution grid topology is more complex than the transmission grid. The characteristics of the distributed grid that makes it difficult to apply methods used in the transmission grid include the following: (i) The area covered by the distribution grid is smaller making small errors in localization much more complicated to correct; (ii) Distribution grids have many branches on any given feeder; (iii) The transmission grid largely focuses on one-way transfer of power from upstream (power generation plants) to downstream (cities). In the distribution grid, there are frequent loops and switches that may cause the flow of power along a section of the network to be reversed or to take a different path entirely; (iv) The structure of the distribution grid is more dynamic since new customers can be added to the distribution grid; (v) Distribution networks have portions of their network placed underground. This changes the properties assumed in the existing work in fault localization for the transmission grid. Figure 3.1 graphically depicts the difference in topology between transmission and distribution grids.

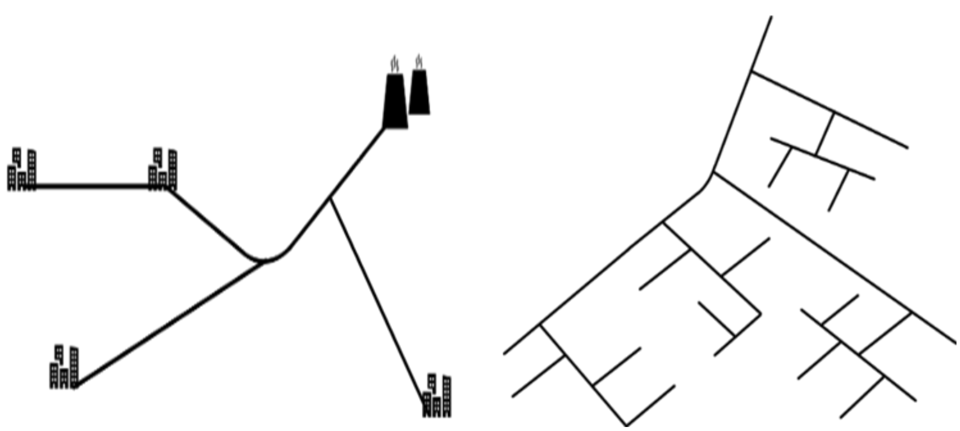

Figure 3.1. Transmission and distribution grid examples

More recently there has been increased use of sensors (e.g. [Trindade et al. 2014], [Trindade and Freitas 2017], [Škumát and Ž 2019], [Sapountzoglou et al. 2019]) to provide more data on the distribution grid. These solutions show promise. However, a challenge with the use of multiple sensor devices is the synchronization of these devices. The fault data recorded by each device must be synchronized so that it can be used effectively. This is especially challenging because faults and the data they generate typically last fractions of seconds. This requires a highly accurate and specialized clock as a few microseconds of error equates to hundreds of meters in distance calculations. The papers that use sensors typically assume that synchronization is already done.

This paper's contribution is a method that allows the synchronization of data collected by devices at multiple locations in the distribution grid without reliance on the accuracy of the clocks used in each device. Faults often manifest themselves as a sudden drop or surge in the current. Before the sudden drop or surge, an incident wave is generated and travels throughout the grid. This is followed by other waves that travel as the result of incident wave reflecting off the ends of the power cables. Sensors can be used to detect the incident wave. With this information and the knowledge of the distance between sensors and propagation speed it is possible to determine the distance to a fault.

The rest of the paper is organized as follows. In section 3.2 different approaches to synchronization are introduced. Section 3.3 provides background information for the environment in which the synchronization method will be applied. Section 3.4 highlights important concepts for the presented method. Section 3.5 provides specific details on the method presented. Section 3.6 looks at the performance of the proposed approach and section 3.7 draws final conclusions.

3.2 Related Work

The introduction discusses the limitations of fault diagnosis approaches in the distribution grid. This section focusses on synchronization. The synchronization of nodes referenced in the literature would be analogous to the sensor devices in our system. Viswanathan et al [Viswanathan et al. 2016] identified two categories of synchronization typically used for sensor-based applications: (i) Exchange of messages between the nodes of the system; (ii) periodic external signals that are not part of the system. The first synchronization category relies on a pre-existing communications network. The external signal methods are of particular interest to our work given the lack of a network and high cost in communication.

3.2.1 Internal Message Synchronization

Internal message synchronization methods are techniques that allow nodes on the network to send messages to each other to enable them to synchronize their clocks. [Elson et al. 2002] present the Reference-Broadcast Synchronization (RBS) scheme. The message sent in this scheme does not provide a timestamp. Instead, the arrival times of the messages are used to synchronize clocks. A two-step protocol is presented in [Ganeriwal et al. 2003]. The first step determines a hierarchy for all nodes and then pairwise synchronization is performed among the nodes, ensuring that all nodes are synchronized to the reference node at the top of the hierarchy. A protocol that is resistant to node failures is presented by [Maroti et al. 2004]. This protocol becomes resistant to node failures by dynamically updating the topology of the node network and periodically flooding the system with synchronization messages. AMAC-layer time stamping, and error compensation using clock skew estimation is used by [Maroti et al. 2004]. The solutions in this category require regular communication among all nodes and thus assumes an existing network between all nodes.

3.2.2 External Signal Synchronization

There are a number of external signals that have been used as time keeping events in order to allow devices that can detect them to synchronize their clocks. Time keeping radio station signals are used by [Chen et al. 2011]; [Li et al. 2011] uses the Radio Data System; [Hao et al. 2011] take advantage of Wi-Fi beacons using Zigbee nodes to detect the beacons; [Rowe et al. 2009] designed a device that utilizes electromagnetic radiation to calibrate clocks by detecting ac cycles; [Li et al. 2012] employs light sensors to detect the flickering of fluorescent lights to calibrate clocks. The solutions that utilize external signals allow devices to operate autonomously. These solutions do not explicitly require an existing communication network as the synchronization signal is not passed through a network. These methods take advantage of signals that already exist in their environment, which allows the system to synchronize without communication. The proposed method also leverages external signals.

3.3 Background: Electrical Power Generation and Faults

3.3.1 Power Generation

Power is generated at power plants. Power plants use generators to move magnets near a wire to create a steady flow of electrons. When a conductor (e.g., copper wire in the shape of coil) is placed within changing magnetic fields, the electrons in the conductor move which means that an electric current is generated. The movement of the magnet is done through a turbine which is a device that burns fuels to rotate the magnets.

Power is generated as alternating current. The magnetic induction causes current to flow relative to the polarity of the end of the magnet. If the end of the magnet next to the coil is positive, then the current flows towards the magnet. If it is the negative end, then the current will flow away from the magnet. As the magnet rotates, and the positive end gets further away from the coil the current weakens until it falls to zero. As the negative end of the magnet moves closer to the coil the current increases again but negatively (negative and positive are used in this context to indicate direction and do not cancel each other out). This effectively creates an alternating current between the maximum current in both directions. When this is plotted over time it creates a sine waveform as seen in figure 3.2. Power is generated in three phases, using 3 coils.

Each coil is separated by 120 degrees and generates its own alternating current as the magnet rotates. Each coil would therefore have power but at different times. Each coil provides a phase where each phase is essentially a feeder (a line used to transmit electrical power). If the current on each coil is plotted over time and placed on the same graph, then the sine waves show what the three-phase alternating current looks like with respect to the change in current over time. This is shown in figure 3.2.

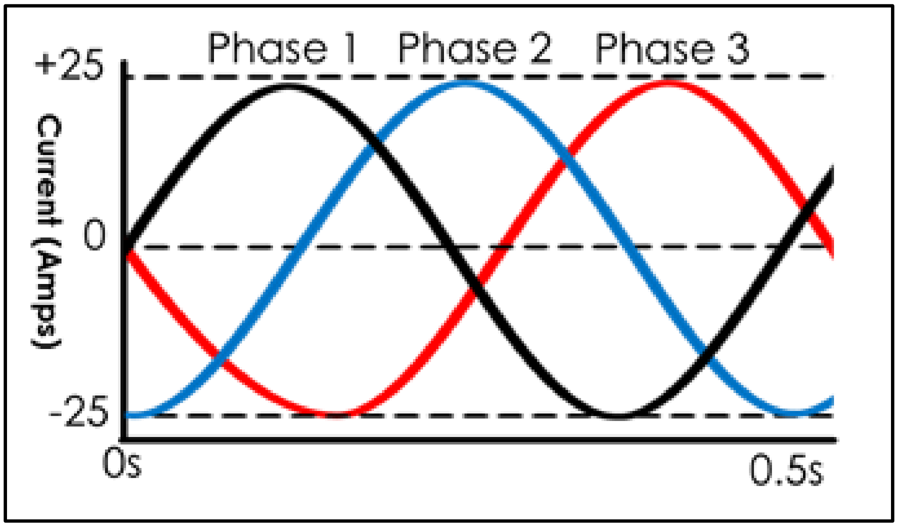

Figure 3.2 Three phase sine waves

The current shown in figure 3.2 is the expected change in current over time (or normal current) created by the oscillation of the magnets towards and away from the coils.

3.3.2 Faults

This section describes the types of faults, the devices used to protect the power grid when a fault occurs and fault detection. In the electrical grid power is usually distributed using three phases. This means that there are three lines used to move power from one location to another.

3.3.2.1 Handling Faults

A fault is detected by reclosers, relays, fuses and breakers. The area that the fault occurs in is turned off by the reclosures and relays. Once the location of the fault is determined and fixed, power can be restored. In figure 3.3, a fault occurs in the area powered by phase 2 (the blue feeder). When the fault is detected by the recloser, it opens and shuts of power to the entire area powered by that phase. There is no knowledge of the location of the fault. It is only known that something has gone wrong and caused the power to be turned off. If this area covers a two- or four-mile area, that entire area has to be visually inspected. This challenge is compounded by the fact that some of these areas of the grid are buried underground.

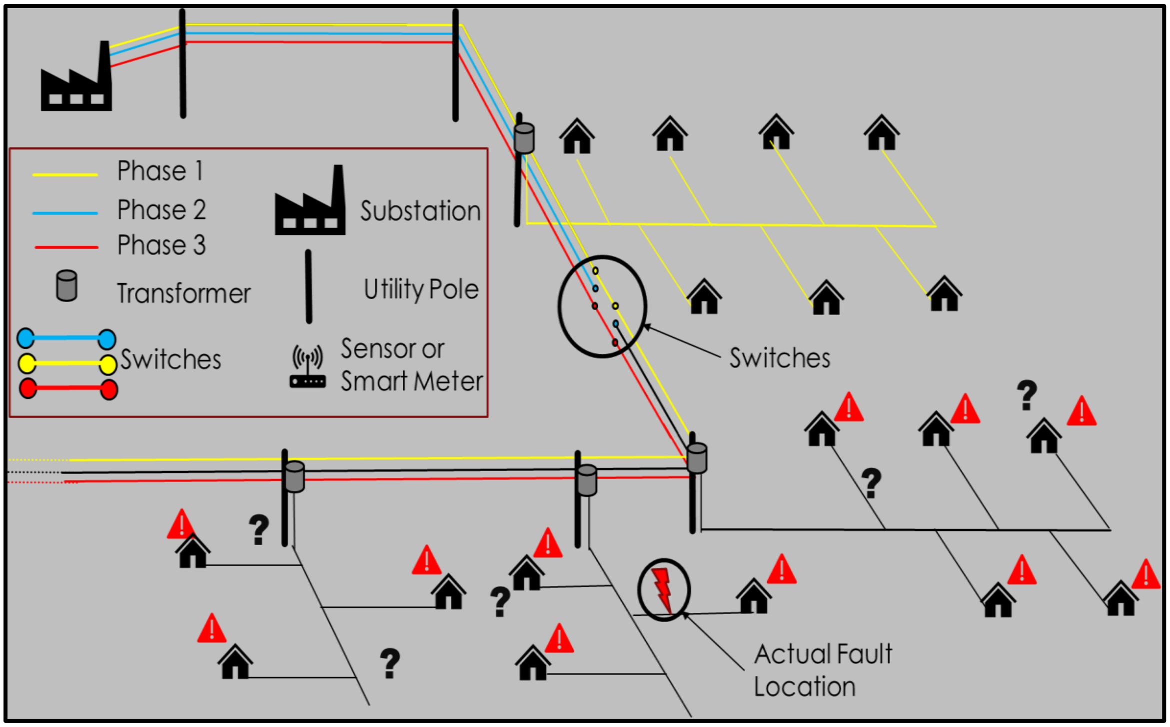

Figure 3.3 Example of impact area for a fault

For a fault to be localized and fixed the presence of a fault must first be detected. With respect to fault detection there are two types of faults: High Impedance Faults (HIFs) and Low Impedance Faults (LIFs). HIFs are more difficult to detect since the fault mimics the normal operation of the grid [Magnago and Abur 1998]. However, only 5 – 20 percent of all faults are HIFs [Hizam and Crossley 2006]. The research conducted therefore focuses on localizing LIFs since it is more common.

A LIF allows large amounts of electricity to flow through it. The electrical grid uses devices that detect LIF currents by monitoring for abnormal currents or voltages present on the line. An abnormal current is identified as an interruption(drop) or surge in the flow of current. Figure 3.4 shows the current before and during a fault on a single phase. These drops and surges that constitute an abnormal current are preceded by a set of travelling waves that propagate through the grid. A fault event (like a downed power line) creates a travelling wave referred to as the *incident wave*. This incident wave travels from the fault point in all directions along the power cable and through the network. This travelling wave is the first peak seen in Figure 3.4. The remaining peaks seen are the result of that incident wave bouncing or reflecting off various components in the electrical grid. These *echo* waves are referred to as reflected waves. The incident wave and reflected waves generated by a fault are referred to as a fault signal.

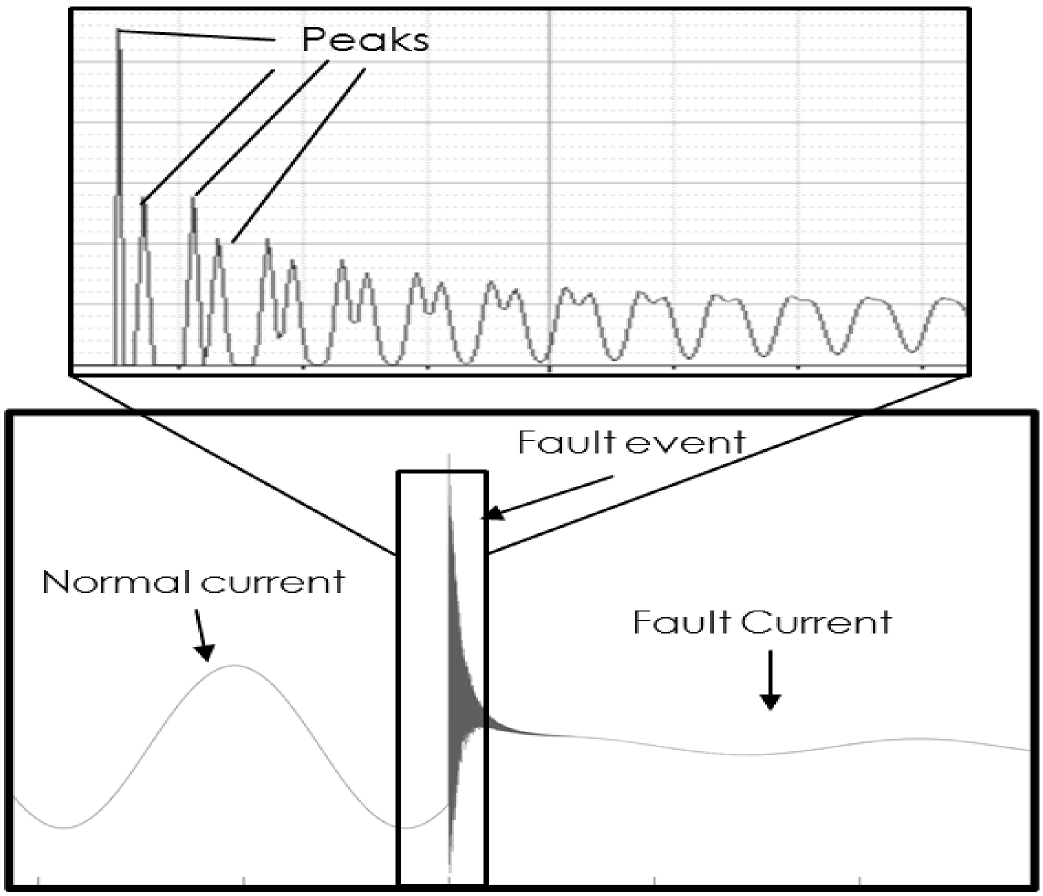

Figure 3.4 Showing simulated current data before, during and after a fault event

3.4 Recloser Signals

The method presented in this work can be categorized as an external signal synchronization method since we use signals generated by components of the electric grid. We assume the use of sensors that detect the signals from these components. In our proposed solution the external signals are the set of travelling waves emitted from the recloser on the electrical grid. In distribution feeders there is a recloser device that is activated whenever a fault is detected. Typically, reclosers are located at the beginning of a feeder and therefore control large distribution areas. The recloser is responsible for clearing(removing) any temporary faults. These are faults that can be cleared by de-energizing the line for a short period of time. An example of a temporary fault is a conductive material making brief contact with a power cable creating an arc through which electrical current flows. De-energizing the power cable will break the electrical arc and end the erroneous flow of electricity. When an erroneous fault current is detected the recloser opens and disconnects power to the feeder to give the temporary fault time to clear (in our example break the flow of electricity through the arc). After a predefined time (usually a few seconds), the recloser then reconnects the feeder. If the fault current is then back within normal range, the recloser leaves the power cable/line connected. If the recloser still detects a fault current it will disconnect the feeder again. The number of times the recloser will attempt to clear the fault is defined by the power company but is typically three times [Industry-Partner 2018] .

The fault itself causes the first set of travelling waves (fault signal). After this is detected, an attempt by the recloser to clear the fault results in the insertion of two more sets of travelling waves (recloser signals) into the feeder. All three sets of travelling waves cause similar fluctuations in current as seen in Figure 3.5. The difference between a fault's set of travelling waves (fault signal) and the recloser's first and second sets of travelling waves (recloser signals) is that the feeder will still have current flowing through it (in the form of abnormal current or fault current) after a fault. However, after the first recloser signal (when opening), the current is reduced to zero and after the second recloser signal (when closing) the flow of current resumes. Figure 3.5 shows the measured current just before and during a fault and the corresponding recloser activation.

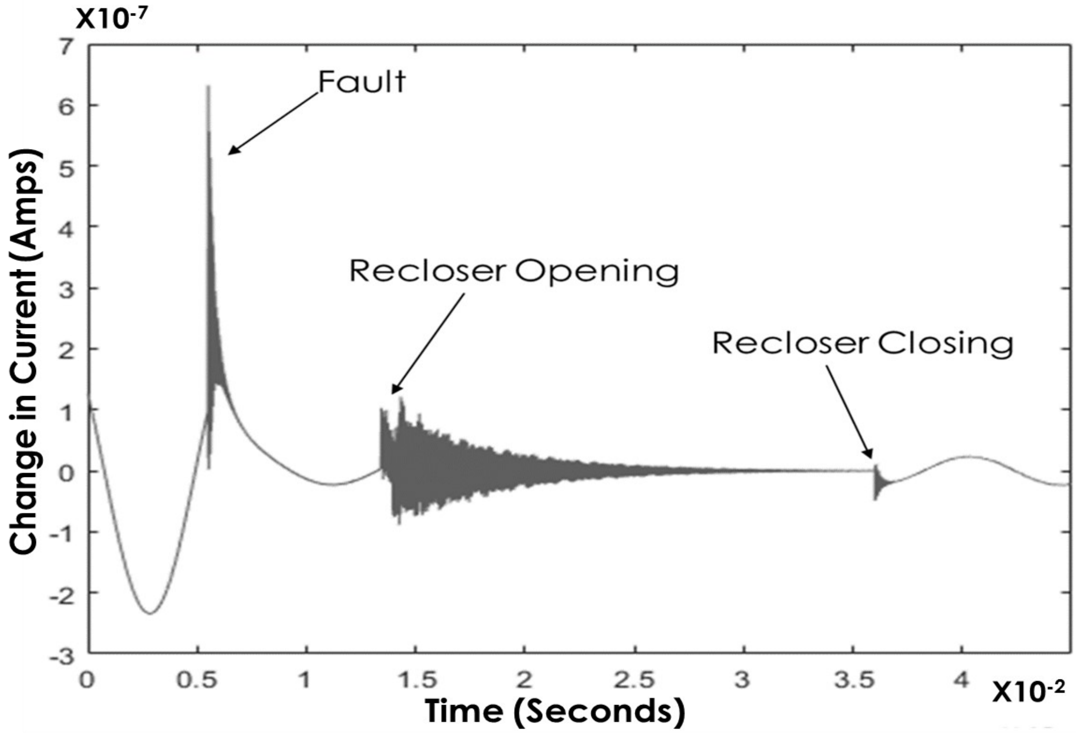

Figure 3.5 Change in current of measured fault and reclosure signal data for one phase

Normal operating current can be observed from 0 to 0.5 seconds. This is the expected sign wave from the natural flow of power from the power generation facility (seen previously in figure 3.2). The effect of the fault is seen at 0.5 seconds. Here we can see that the current suddenly changes significantly but some power is still flowing (fault current) up until 1.3 seconds. At 1.3 seconds the recloser activates (by opening and halting the flow of power) and de-energizes the feeder creating another disturbance in electrical current. The recloser opening causes the abnormal fault current to drop to zero (the first recloser signal). After the line is de-energized, the recloser reconnects the power at 3.52 seconds and a surge in current is then detected as power resumes its flow (the second recloser signal). In this case the current is still abnormal (reduced in this case) and the recloser will repeat the process or remain open and permanently disconnect the power.

3.5 Using Reclosers for Synchronization and Fault Localization

This section describes how we leverage the recloser incident wave to synchronize the incident wave from the fault detected by the devices.

The locations of sensors and reclosures on the grid are known and hence the distances between them are known. The propagation speed of waves on the given line is also known. With this information the time it takes for the recloser's travelling wave to reach some sensor A can be calculated using equation 1.

$$T_A^R = \text{dist}(R, A) * s \qquad (1)$$

where *dist(R,A)* is the distance between the recloser and sensor A, and *s* denotes the propagation speed of waves on the given line. T_A^R is the time that sensor A receives the reclosure signal while T_B^R is the time that sensor B receives the reclosure signal. The relationship between distance and delays allows us to measure delays and then convert delay values to distances. The delay between the reclosure signal arriving at sensor A and B is represented by $\text{delay}_{A,B}^R$. Thus, the delay value represents the difference between the times that sensors A and B detect the reclosure signal. Since the locations of sensor A and B are known as well as the location of the recloser and the propagation speed of line, the delay between the arrival times of the recloser to each sensor can be determined aprori. This delay between the arrival times of the recloser to each sensor can be calculated by equation 2.

$$\text{delay}_{A,B}^R = T_A^R - T_B^R \qquad (2)$$

The measured delay between the fault and recloser signals detected by sensor *x* is represented by $\text{delay}_f^R(x)$. Equation 3 is used to calculate the time delay between the recloser wave and the fault wave measured by a single sensor, x.

$$\text{delay}_f^R(x) = T_x^R - T_x^f \qquad (3)$$

By taking the delay between T_x^R and T_x^f the arrival of the fault signal at sensor *x* can be measured with respect to the recloser signal rather than the possibly incorrect time of sensor *x*'s clock. The expected delay of a signal travelling from sensor A to sensor B is represented by $\mathrm{delay}_{\mathrm{B}}^{\mathrm{A}}$. The $\mathrm{delay}_{\mathrm{B}}^{\mathrm{A}}$, $\mathrm{delay}_{\mathrm{f}}^{\mathrm{R}}(\mathrm{A})$, $\mathrm{delay}_{\mathrm{f}}^{\mathrm{R}}(\mathrm{B})$ and $\mathrm{delay}_{\mathrm{A,B}}^{\mathrm{R}}$ can be used to factor out the recloser times and provide a way to localize a fault *f* from sensor A using equation 4. The function abs() takes the absolute value of the calculation and *s* is the propagation speed of waves on the given line.

$$dist(A,f) = \mathrm{abs}\left(\frac{\mathrm{delay}_{\mathrm{B}}^{\mathrm{A}} - \left((\mathrm{delay}_{\mathrm{f}}^{\mathrm{R}\prime}(\mathrm{B}) - delay_{A,B}^{R}) - \mathrm{delay}_{\mathrm{f}}^{\mathrm{R}}(\mathrm{A})\right)}{2}\right) / \mathrm{s} \quad (4)$$

The delay components which make up the equation for fault localization (equation 1) only use the local times recorded by the sensor for which that delay is being determined. The result of using the delays between fault and recloser signals detected by each sensor ensures that fault location calculations do not require a global clock. The testing of this equation is presented in section 3.6.

3.6 Evaluation

To determine the effectiveness of this method we simulated a distribution line with two sensors as shown in figure 3.6. One end is connected to a substation providing power, which then flows along the line through the recloser and continues to a customer.

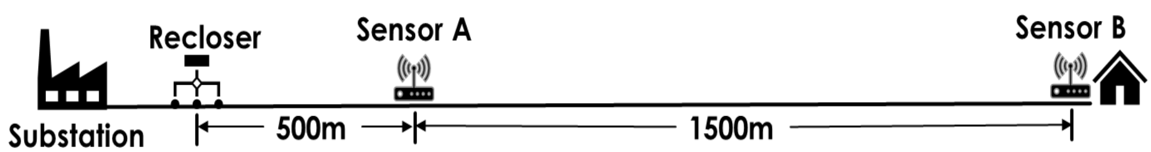

Figure 3.6 Example of a cable sensors and recloser

Using the known distances, the propagation delay of the line and equation 1, faults occurring between the two sensors A and B can be localized. Genuine testing for this solution's application in a real scenario, required the simulation of an actual fault and applying the solution in this scenario to determine its effectiveness. For this testing

Simulink is used to create a model of a distribution line using figure 3.6 as the template. The recloser and sensors are added, and a fault is then simulated between the two sensors.

3.6.1 Establishing the Sample Rate

For the proof of concept in a real-world scenario, a sample rate of 5Mhz was chosen. This initial value was chosen as a 5Mhz sample rate is one of the less expensive devices that can be used for recording the fault data needed for this solution. This sample rate determines how often the electrical current is measured at each sensor's location. The expected timing of the reclosure and fault wave arrival times was calculated and then compared with the arrival times recorded during the simulation. This was done for eight measurements of recloser and fault waves. Eight tests allowed two faults to be simulated on the rising and falling segments of positive and negative current values generated by alternating current (as explained using figure 3.2). The measurement of arrival times of the recloser wave at 5Mhz resulted in an average error of 57.64 meters to be introduced into the calculation of the fault location. The error in the measurement of arrival times of the fault incident wave at 5Mhz resulted in an average error of 70.17 meters to be introduced into the calculation of the fault location.

These errors in recloser and fault wave measurements due to a 5Mhz sample rate resulted in errors of final fault location calculations for the tests to be from 23 meters up to 168 meters. With a sample rate of 5Mhz there is a gap of 2e-7 seconds (0.2 microseconds) between samples. This means that the current change that is indicative of an incident wave could be detected up to 2e-7 seconds after its arrival. This is because in a worst-case scenario, if the sample is taken just before the current changes, the entire 2e-7 seconds would pass before the next sample is taken and the change in current is detected. The impact of such a delay can be seen when using the signal propagation delay on a given cable. The propagation delay of the cable in our scenario is 288×10^6 meters per second. When converted to distance that 2e-7 seconds (0.2 microseconds) translates to 57 meters. This means that at 5Mhz every sample can have an error of 57 meters. As two samples are taken, 114 meters of error can be introduced due to the sample rate for each sensor.

At this point it was noted that a correct sample rate must first be determined and then the solution can be tested. The optimal sample rate is dependent on the signal speed of the cable and the desired error range. More practically, this would also include the cost of the device, as devices with higher sample rates are more expensive.

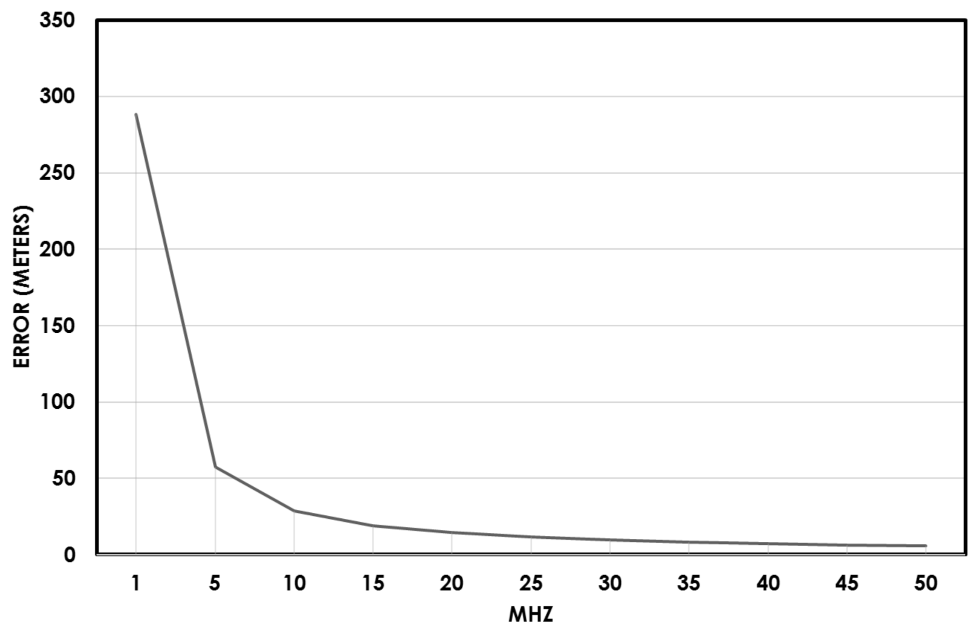

Figure 3.7 Showing error rate reduction from increase in sample rate

Figure 3.7 shows the decrease in error achieved by the increase in sample rate. In figure 3.7 the improvement of error for sample rate increase quickly diminishes requiring large increases of sampling rate for a small reduction in error.

In a real-world implementation, the sample rate chosen may vary based on the signal speed of the cable used and the cost of the devices. In our tests 20Mhz is chosen because an error of 14 meters provides a high level of accuracy for the computational and storage cost of running a simulation at that sample rate. When implementing the solution, larger sample rates can be used to reduce the error entirely if needed. To remove the error entirely the chosen sample rate should be equal to the propagation speed of the power cables used in the grid being monitored.

3.6.2 Final Evaluation

Using 20MHz as the sampling rate should reduce the error from 0.2 microseconds to 0.05 microseconds (or 50 nano seconds) per sample for a signal propagation speed of 288×10^6 providing an overall reduction in fault location error to under 30 meters. This would then show that it is possible to control (reduce or remove) the error in fault localization by adjusting the sampling rate according to the propagation speed of the signals on the electrical cable being monitored.

Ten tests were conducted with a sample rate of 20 Megahertz to confirm that the error falls withing the expected range for such a sample rate given a signal propagation speed of 288×10^6. Random fault locations were placed along a cable and simulated in Simulink. The delay between the fault signal arriving at each sensor was calculated using the recloser to synchronize. The delay was then used to determine the fault location. The results are shown in table 3.1. The average error of these test was 15.73 meters with a maximum error of 26.51 meters.

Table 3.1. Test Results at 20MHz

Measured Distance (meters)	Actual Distance (meters)	Error (meters)
1289.63	1300.00	10.37
1181.56	1200.00	18.44
2377.52	2400.00	22.48
1887.61	1900.00	12.39
1988.47	2000.00	11.53
2074.93	2100.00	25.07
987.03	1000.00	12.97
489.91	500.00	10.09
792.51	800.00	7.49
3573.49	3600.00	26.51

3.7 Synchronization of Multiple Sensors

The example used in Section 3.6 uses two sensors to show how the recloser can be used to synchronize the fault signal's arrival time at each sensor and determine the location of the fault in a single equation. An important element in fault localization with this equation is that even though there may be many sensors in the distribution grid, only *two* sensors are needed to determine a fault's location using Equation 4 from section 3.5. The only requirement in choosing the two sensors is that the sensors must be on either side of the fault (the fault must be between the two sensors chosen).

It is also possible to use the recloser signal to synchronize the data separately from the localization of the fault. The steps to synchronize the travelling wave fault data across multiple sensors are outlined below:

- Step 1 - Use the known distances between the sensors and the recloser to calculate the expected delays between the recloser signal arrival time at each sensor.
- Step 2 - Calculate the delay between the fault and recloser signals for each sensor.
- Step 3 - Use the expected delays of the recloser signals between sensors (calculated in step 1) to correctly align the recloser signals and align the fault signals relative to their corresponding recloser signals (using the delays calculated in step 2).

Figure 3.8 shows an example of a grid with a recloser, three sensors and a fault occurring at the fault point on the grid. Table 3.2 shows the corresponding adjacency matrix of Figure 3.8 for the sensors and recloser.

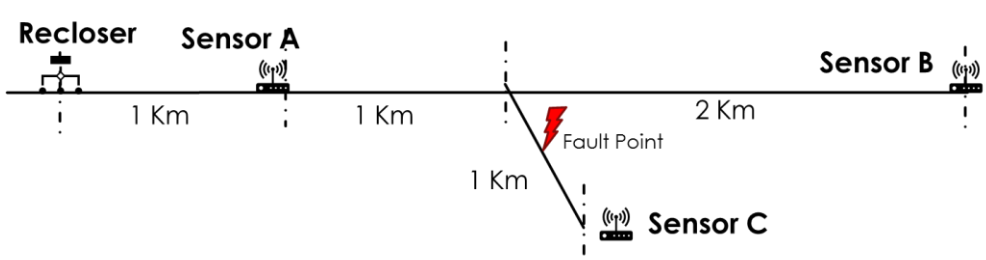

Figure 3.8 An example grid using three devices/sensors

Table 3.2 Adjacency matrix of sensors and reclosers

	Recloser (Km)	Sensor A (Km)	Sensor B (Km)	Sensor C (Km)
Recloser (Km)	0	1	4	3
Sensor A (Km)	1	0	3	2
Sensor B (Km)	4	3	0	3
Sensor C (Km)	3	2	3	0

Table 3.3 Delays between recloser signal times of each sensor

Recloser Delay		
A - B	A - C	B - C
-0.03	-0.02	0.01

The first step outlined above is completed using a propagation delay of 0.01s per Km, the delays of the recloser signal between each sensor is calculated and shown in Table 3.3. The second step, calculating the delays between the fault and recloser signal is shown in Figure 3.9. The delays for sensor A, B and C are 0.02, 0.04 and 0.05 respectively as shown in Table 3.4.

Table 3.4 Delays between recloser and fault signals for each sensor

Recloser - Fault Delay		
A	B	C
0.02	0.04	0.05

The values in Table 3.4 are generated by simulating the example shown in Figure 3.8 and recording the recloser and fault signals at each sensor. Using Sensor A as an example, say the fault occurs at T = 0. Sensor A will get the fault signal at T = 0.015. That fault signal will reach the recloser at T = 0.025. The recloser signal would then leave the recloser and arrive at sensor A at T = 0.035. Using Equation 3 from Section 3.5 the Recloser – Fault Delay for Sensor A would be calculated as shown below.

$$\text{delay}_{\text{f}}^{\text{R}}(\text{A}) = \text{T}_{A}^{R} - \text{T}_{A}^{f} = 0.035 - 0.015 = 0.02$$

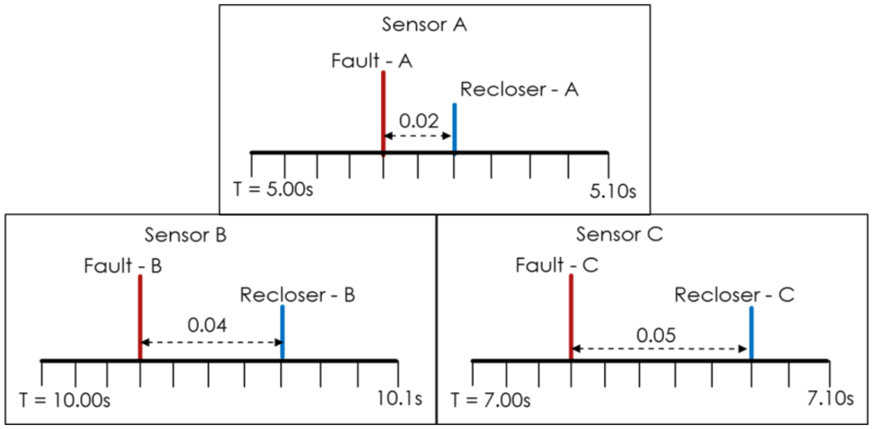

Figure 3.9 Showing the measured delay between fault and recloser signals of each sensor

The final step is to place the recloser signals the appropriate distance apart using the recloser delays between each sensor from Table 3.3. The fault signal for each sensor can then be inserted at the recorded distance from the recloser signals for each sensor using the Reclosure-Fault delays from Table 3.4. Figure 3.10 shows the final synchronized fault and recloser signals.

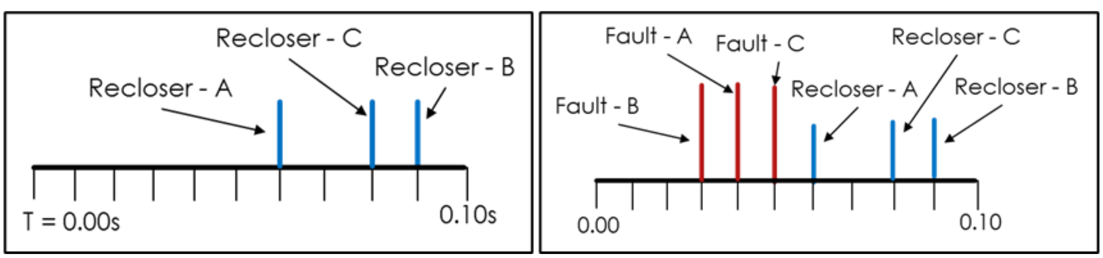

Figure 3.10 The alignment of the recloser signals (left) and then the fault signals (right) across multiple devices.

Using equation 4 from Section 3.5 the fault location can be determined by using sensor A and sensor C's data and is shown below.

$$\mathrm{abs}\left(\frac{\mathrm{delay}_{\mathrm{C}}^{\mathrm{A}} - \left((\mathrm{delay}_{\mathrm{f}}^{\mathrm{R'}}(\mathrm{C}) - delay_{A,C}^{R}) - \mathrm{delay}_{\mathrm{f}}^{\mathrm{R}}(\mathrm{A})\right)}{2}\right)/\mathrm{s}$$

$$\text{abs}\left(\frac{0.02-(0.05-(-0.02))-0.02)}{2}\right)/s = \text{abs}\left(\frac{-0.03}{2}\right)/0.01 = 1.5\text{Km}$$

The final calculation is that the fault is 1.5Km away from sensor A. In this section the synchronization of multiple devices is shown separately from the calculation of the fault location. It provides an example of how all sensors in a grid can be synchronized. Then using equation 4 defined in Section 3.5, the fault is localized which only requires the use of two sensors on either side of the fault.

3.8 Conclusion and Future Work

In this paper we present a synchronization method that shows much promise in allowing faults to be localized by using more than one device. This would enhance the stability of the localization system as complete reliance is no longer placed on the accuracy of a single measurement from a single device. Furthermore, the use of expensive clocks or components to maintain synchronization of the devices and their data is not needed, providing a simpler and cheaper system. The purpose of these tests was to determine if equation 1 can synchronize the data collected by the sensors without the synchronization of their clocks and localize faults. These results show that the solution represented in equation 1 can synchronize the data collected by sensors without the need for synchronization between the clocks of each of the sensors. The method was shown to successfully synchronize the data and localize the faults. However, using an arbitrary sampling frequency of 5Mhz resulted in a small error of up to 0.4 microseconds in the final fault location. The remaining testing focused on the study of these small errors and determining a solution for reducing or removing these small errors. The errors were found to be occurring due to the sampling rate used to collect the data and the propagation speed of the signals being recorded in a real-world scenario and not a part of the synchronization solution. The final tests showed that adjusting the sampling rate when considering the propagation speed of the signals measured allowed the error in fault localization to be controlled. These tests show that the synchronization solution can successfully synchronize fault data and localize faults not only in theory but in a real-world scenario. The only present error external to the equation and is due to the propagation speed of the signal and the sample rate used to record the data. Regarding the external factors generating this error, the test also showed that this error can be removed by increasing the sample rate used to record the data. Once the sample rate is adjusted according to the propagation speed of the signals, the accuracy of the data and the accuracy of the fault localization can be assured.

The next steps for this work involve developing procedures for identifying devices that have errors in the measured data itself by leveraging redundancies that can be placed in a synchronized monitoring system. In relation to the application of this work towards the goal of fault localization, the next step is to develop the multi device fault localization method. The synchronization method presented in this paper coupled with a fault localization method, will offer a cheaper and more complete solution for fault localization in distribution grids.

3.9 Summary

Chapter 3 is dedicated to taking the primary step in the development of a more effective fault localization solution for distribution grids. Chapter 3 presents a significant contribution to fault localization research by providing a method that synchronizes the travelling wave data recorded from multiple devices for fault localization. At the time of the writing of this dissertation, there is no method that can synchronize the travelling wave data recorded by multiple devices independently of an external synchronization system like GPS. The benefit of establishing this synchronization method is that it satisfies the need for travelling wave data synchronization, without incurring the cost of adding systems like GPS to the devices. This synchronization method is shown in an equation that allows a fault occurring between two sensors/devices to be localized. Chapter 4 generalizes this method so that it can be applied in a complex distribution grid with multiple branches and sensors.

The existing solutions have limited success in addressing the main challenge with fault localization methods in the distribution grid. The challenge is that multiple possible locations are generated for the fault rather than just one location when localizing a fault in the distribution grid. Chapter 4 looks at this challenge and takes a unique approach at localizing faults using travelling waves. The fault localization solution presented in Chapter 4 splits the distribution grid into segments and generates unique identifiers for each segment to create an effective means by which only a single location is generated for the fault.

3.10 Bibliography

Chen, Y., Q. Wang, M. Chang and A. Terzis. Ultra-low power time synchronization using passive radio receivers. *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, (2011).

Crossley, P. A. and P. G. McLaren. "Distance Protection Based on Travelling Waves." *IEEE Transactions on Power Apparatus and Systems* **PAS-102**(9): 2971-2983, (1983).

Das, S., S. Kulkarni, N. Karnik and S. Santoso. Distribution fault location using short-circuit fault current profile approach. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Elson, J., L. Girod and D. Estrin. "Fine-grained network time synchronization using reference broadcasts." *SIGOPS Oper. Syst. Rev.* **36**(SI): 147-163, (2002).

Ganeriwal, S., R. Kumar and M. B. Srivastava. Timing-sync protocol for sensor networks, ACM**:** 138-149, (2003).

Hao, T., R. Zhou, G. Xing and M. Mutka. WizSync: Exploiting Wi-Fi Infrastructure for Clock Synchronization in Wireless Sensor Networks. *2011 IEEE 32nd Real-Time Systems Symposium*, (2011).

Hizam, H. and P. A. Crossley. Single Ended Fault Location Technique On a Radial Distribution Network Using Fault Generated Current Signals, (2006).

Industry-Partner. Ontario Electricity Provider, (2018).

Karnik, N., S. Das, S. Kulkarni and S. Santoso. Effect of load current on fault location estimates of impedance-based methods. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Li, L., G. Xing, L. Sun, W. Huangfu, R. Zhou and H. Zhu. Exploiting FM radio data system for adaptive clock calibration in sensor networks, ACM**:** 169-182, (2011).

Li, Z., C. Li, W. Chen, J. Dai, M. Li, X.-Y. Li and Y. Liu. Clock calibration using fluorescent lighting, ACM**:** 463-466, (2012).

Magnago, F. H. and A. Abur. "Fault location using wavelets." *IEEE Transactions on Power Delivery* **13**(4): 1475-1480, (1998).

Marguet, R. "Improved fault localization method for electrical power distribution networks." (2015).

Maroti, M., B. Kusy, G. Simon and A. Ledeczi. The flooding time synchronization protocol, ACM**:** 39-49, (2004).

Rowe, A., V. Gupta and R. Rajkumar. Low-power clock synchronization using electromagnetic energy radiating from AC power lines, ACM**:** 211-224, (2009).

Saha, M., F. Provoost and E. Rosolowski. "Fault location method for MV cable network." (2001).

Sapountzoglou, N., B. Raison and N. Silva. Fault Detection and Localization in LV Smart Grids. *2019 IEEE Milan PowerTech*, (2019).

Škumát, T. and E. Ž. Open Phase Fault Localization in a Medium Voltage Distribution Grid Using Smart Meters. *2019 20th International Scientific Conference on Electric Power Engineering (EPE)*, (2019).

Takagi, T., Y. Yamakoshi, M. Yamaura, R. Kondow and T. Matsushima. "Development of a New Type Fault Locator Using the One-Terminal Voltage and Current Data." *IEEE Transactions on Power Apparatus and Systems* **PAS-101**(8): 2892-2898, (1982).

Trindade, F. C. L. and W. Freitas. "Low Voltage Zones to Support Fault Location in Distribution Systems With Smart Meters." *IEEE Transactions on Smart Grid* **8**(6): 2765-2774, (2017).

Trindade, F. C. L., W. Freitas and J. C. M. Vieira. "Fault Location in Distribution Systems Based on Smart Feeder Meters." *IEEE Transactions on Power Delivery* **29**(1): 251-260, (2014).

Viswanathan, S., R. Tan and D. K. Y. Yau. Exploiting Power Grid for Accurate and Secure Clock Synchronization in Industrial IoT. *2016 IEEE Real-Time Systems Symposium (RTSS)*, (2016).

Chapter 4

4 Fault Localization in Smart Grids Using Segmentation

4.1 Introduction

Chapter 3 presents a synchronization method for devices that record travelling waves. It also shows how faults can be localized on a small example grid, using this synchronized travelling wave data. This chapter (Chapter 4) now builds on that initial work and creates a complete fault localization solution that can be applied to a larger, complete distribution grid.

The electrical grid consists of power plants, the transmission grid and the distribution grid. Power plants traditionally generate power from either coal, water, oil or nuclear. Power plants are typically located miles away from power consumers. In order to transmit the power over these large distances to consumers the power is sent over the transmission grid through a transmission substation that converts the power to a high voltage suitable for long distance transmission. Once it reaches a cluster of consumers the distribution grid is used to deliver the power to the consumers. The transition from the transmission grid to the distribution grid is through a distribution substation which uses transformers to reduce the power voltage before it is transmitted to the consumers.

Should a fault occur, the grid is designed to quickly stop sending power to that section of the grid. This protects people, the electrical grid, and its equipment. Specialized equipment like relays, fuses and switches allow the power to be shut off quickly. The restoration of power requires not only the detection of the presence of the fault but the location of the fault. Determining the location of the fault is challenging since a fault that occurs anywhere in the distribution grid causes the area around the fault to be shut down.

There are a number of solutions that have been developed for both the transmission and distribution grids [Marguet 2015]. However, much of the existing work on fault localization for the distribution grid typically applies variations of existing techniques used in the transmission grid and thus does not always consider the differences in topology. As seen in Figure 4.1, the distribution and transmission grids differ not only by their voltage level but also by their topologies. The topology of a distribution grid is typically a radial network with a tree-like structure, where the root represents a substation and there is a very high branching factor compared to the transmission grid. The distribution grid also differs from the transmission grid in that it has more components, the information available for the distribution grid typically has more errors as its components change more frequently, and the distribution grid has more conductor types as several transformations are made on the voltage to step it down for consumer usage needs.

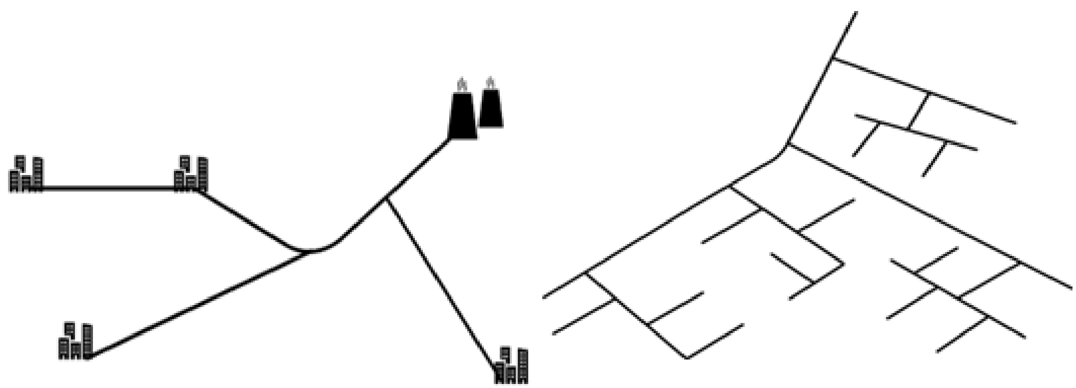

Figure 4.1 Transmission vs Distribution Grid Topology

Transmission power lines are typically from power plants to distribution points with relatively little branching. This makes it easier to perform calculations used to determine fault locations as very few components need to be represented. There are fewer inaccuracies in the data representing the transmission grid's components as it is not altered as often when compared to the distribution grid, which makes it easier to develop more accurate localization methods for the transmission grid. The typical approach for finding a fault in the distribution grid is to send crews to the affected area to find the fault.

This paper describes an approach to fault localization that considers the specific characteristics of the distribution grid. These characteristics are derived from the increased complexity of the distribution grid due to a higher branching factor and more components and voltage changes. The rest of the paper is organized as follows. In section 4.2 different approaches to fault localization in distributed grids are presented. Section 4.3 describes the challenges faced by existing methods. Section 4.4 describes the approach taken by this paper to address these challenges to localize faults. Section 4.5 presents our experiments and results. Section 4.6 draws conclusions and sheds light on future work.

4.2 Related Work

Fault localization approaches for distribution grids can be categorized as impedance-based or as travelling waves [Marguet 2015].

4.2.1 Impedance-Based Methods

Impedance-based methods are described by [Andrade and Leão 2012] to be a calculation of distance that uses the loop that is created between the fault and the point on the line that is being measured. Impedance-based methods [Marguet 2015] include positive reactance ([Saha et al. 2001, Das et al. 2011, Karnik et al. 2011]), loop reactance [Qin et al. 1998] and the Takagi equation [Girgis et al. 1993].

There is existing work on the application of impedance-based methods in the distribution grid (e.g. [Girgis et al. 1993], [Jun et al. 1997], [Seung-Jae et al. 2004], [Gabr et al. 2017]). However, [Mwifunyi et al. 2019] note that these impedance methods use constant impedance load modelling which is somewhat detached from the variability of the load in a real distribution grid, and may provide multiple estimations of possible fault locations.

4.2.2 Travelling Wave Methods

When a fault occurs in the electrical grid an *incident wave* is emitted from the fault location and travels across the grid and reflects off the end of a line and therefore travels back up the line. The *reflected waves* also bounce off the fault location. A sensor can be used to record the arrival times of the waves on a line. This means that for a fault there is one incident wave and multiple reflection waves. The incident wave always arrives at a sensor before the reflection wave. A sensor does not cause a reflection. Traveling wave methods include relay measurements and wavelet transformation [Marguet 2015]. Relay measurements (e.g. [Crossley and McLaren 1983]) are used to detect fault waves and calculate distances using a given propagation delay. Wavelet transform methods are similar to relay measurements but perform a wavelet transform on the signal to extract the surges(waves) and then distances are calculated using a given propagation delay (e.g., [Magnago and Abur 1998]).

Several solutions examine more sophisticated analysis of the travelling waves and add new components to improve accuracy of the localization results. These techniques employ neural networks and genetic algorithms [Pourahmadi-Nakhli and Safavi 2011], [Wen and Chang 1998, Srinivasan et al. 2000]].

4.3 Challenges with Using Existing Methods in the Distribution Grid

This section describes the challenges with applying impedance based and travelling wave methods used in the transmission grid to the distribution grid.

4.3.1 Impedance-Based Method Challenges

Impedance based methods generate a distance from a sensor's location to the fault [Andrade and Leão 2012]. In a transmission grid there is only one power line being monitored and therefore only one location for the fault. A distribution grid however has multiple power lines branching away from the main power line. This results in the distance to the fault creating multiple locations along these branches rather than just one. Example 1 and figure 4.2 illustrate this possibility.

Example 1: Figure 4.2 shows a distribution grid that uses a single sensor at point A. Assume a fault was calculated to be 950 meters away from the sensor. In this example, there are five possible fault locations (shown as red symbols in figure 4.2). As the number of branches and the area monitored increases the more possible fault locations there might be.

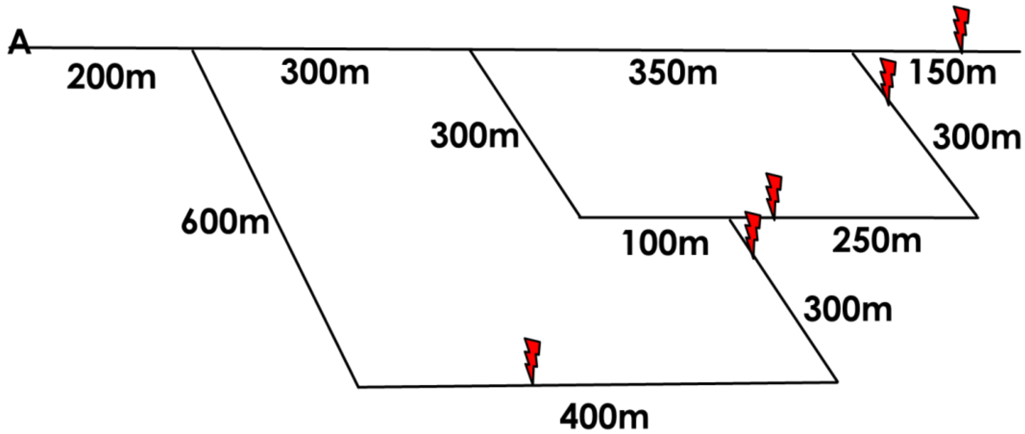

Figure 4.2 Multiple fault locations due to multiple paths

4.3.2 Travelling Wave Method Challenges

Travelling wave methods used in the transmission grid generate a distance from a sensor (seen as S_1 in figure 4.3) using the incident and reflected wave to determine the fault location. Since a transmission grid has one power line there is only one source for the reflected waves which is at the endpoint representing the end of the power line. In distribution grids however, there are many endpoints (referred to as EP_1, EP_2 and EP_3 in figure 4.3). The reflected waves detected can come from any of these endpoints. As a result, a single endpoint and reflected wave cannot be established.

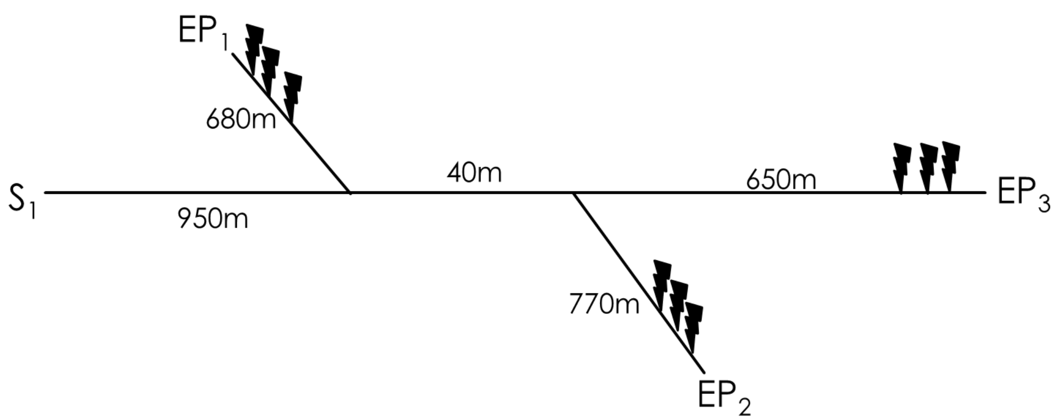

Figure 4.3 Multiple fault locations due to unknown wave sources

The source endpoint of the reflected wave used to localize a fault must be known. If the sources of the reflected waves cannot be determined as is often the case in the distribution grid, then only a list of possible locations will be generated. This is done by using *all* combinations of reflected waves and endpoints. Figure 4.3 shows the nine possible fault locations when the incident wave and just three reflected waves are recorded by the sensor with more than one endpoint. Only the list of possible locations can be determined because we do not know which endpoint caused which wave. All that can be done is to assume all reflected waves came from all endpoints. Existing methods focus on reducing the number of possible fault locations generated by these solutions in the distribution grid.

4.4 Fault Localization

To address the problem of the possibility of multiple fault locations as seen in figure 4.3 we use multi-sensor travelling wave localization and grid segmentation.

4.4.1 Multi-Sensor Travelling Wave Localization

Our approach uses multiple sensors and travelling waves to localize a fault. Rather than using the delay between the incident wave and reflected wave recorded at one sensor, we use multiple sensors and we use the delay between the arrival times of the incident wave for each pair of sensors. In this subsection, we will describe the approach for a single line and in subsequent subsections we will discuss how our approach applies when there are branches.

If sensor *i* detects the incident wave at time t_i and sensor *j* detects the incident wave at t_j, then *delay(i,j)* is calculated as $t_i - t_j$,. The value of *delay(i,j)* represents the difference between the times that sensors *i* and *j* detect the fault signal. If *delay(i,j)* is zero then this indicates that the fault is in the middle of the power line. If *delay(i,j)* is positive then the location of the fault is between the midpoint and sensor *j* and if *delay(i,j)* is negative then the location is between sensor *i* and the midpoint of the power line. The delay value can be converted to a distance from the midpoint of the power line using the distance between *i* and *j*, and the speed, *s*, that fault signals traverse the power line. Using this information, Equation 1 calculates the distance of the fault from sensor *i.* For a single line this is sufficient to determine the fault's location.

$$\text{distance}(i) = \frac{\left(d_{i,j} + (\text{delay}(i,j) * s)\right)}{2} \quad (1)$$

The distance from sensor *j* can be calculated by using *delay(j,i)* in equation 1. Example 2 describes how equation 1 can be used to calculate the distance of the fault from a sensor for a single line.

Example 2: Figure 4.4 shows a power line of length 1000m that uses a fault signal speed of 100m/s. Sensors are placed at each end (represented by *A* and *B*) of the power line in figure 4.4. If the fault occurs at point *A*, the sensor at point *B* detects this 10 seconds after the sensor at point *A*. The delay is calculated as -10 and so the location is calculated to be at point *A*. If the fault occurs at the midpoint then the delay is zero and hence the fault location is 500 meters from the sensor at point *A*.

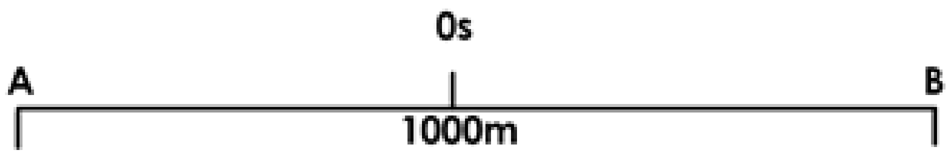

Figure 4.4 Single power line fault localization

With branching, it is possible for there to be multiple locations that are the same distance from a sensor. Segmentation is used to address challenges posed by branching.

4.4.2 Segmentation of Distribution Grid

The distribution grid can be conceptualized as a graph, G = (V, E). Each edge e ∈ E represents a segment of a power line. The set V represents the vertices of the edges. Each edge *e* is associated with a set, *W*, of time windows. Each time window is a pair of values, ($T^l_{i,j}$ (e), $T^u_{i,j}$ (e)), that is associated with a pair of sensors, *i* and *j,* and represents the range of delays for a segment. $T^l_{i,j}$(e) represents the lower bound of the time window for the pair i, j for an edge *e*. $T^u_{i,j}$(e) represents the upper bound of the time window for the pair i, j for an edge *e*. Given the number of sensors, *n*, the number of time windows per edge is *n(n-*

1)/2. Segmentation is used to identify specific parts of the distribution grid to be examined to determine the fault location.

Time windows can be applied to any segment of a power line. Assuming we have sensors, *i* and *j*, and edge (*x,y*), the amount of time before sensor *i* receives a fault signal that occurs at point *x is dist(i,x)/s* and for sensor *j* it is *dist(j,x)/s* and thus the lower and upper limits of the delay for sensors *i* and *j* is defined in Equation 2.

$$T^{l}_{i,j}, T^{u}_{i,j} = \frac{dist(i,x) - dist(j,x)}{s}, \frac{dist(i,y) - dist(j,y)}{s} \quad (2)$$

Example 3 shows how segmentation and time windows can help localize faults.

Example 3: This example shows the concept of segmentation with two sensors placed at points A and B (each end of a power line). The locations of the sensors do not need to be at the end of a power line but can be anywhere on the electrical cable in the grid. It shows how a time-window for a segment of the line between point X and point Y is calculated. Using the distances in figure 4.5 and equation 2, the lower and upper bounds of the time window for the segment between point X and point Y is calculated to be -8.5 and 11.5 respectively.

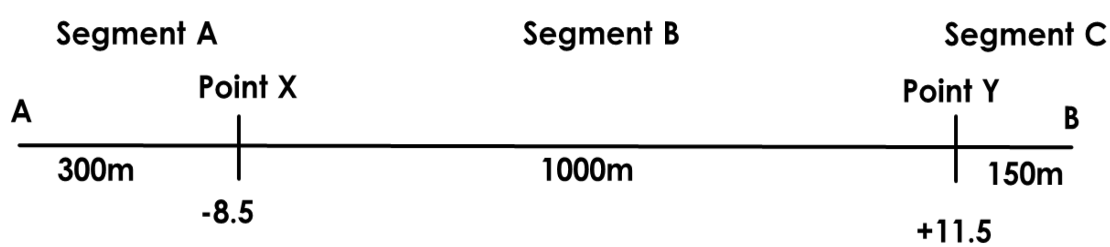

Figure 4.5 Example of power line for time-window calculation

4.4.3 Application of Multi-Sensor and Segmentation Concepts in Fault Localization

This section shows how the use of time-windows and multiple sensors can be used to determine a single location for faults rather than a list of possible locations. We assume that the following information is available: the lengths of the power lines in the grid, the speed that the signal travels along the power lines, locations of the sensors used to detect the fault signal and locations of the edges. Example 4 describes the application of segmentation and time windows for a specific example grid.

Example 4: Figure 4.6 shows an example of the segmentation of a simple distribution grid that uses three sensors denoted by *A, B* and *C*. The grid is segmented into edges. The terms edge and segment are therefore used interchangeably. Example edges include (A,2), (3,4), (3,7), (2,C). With three sensors there are three time-windows for each edge. Each edges' time windows are referred to as an edge key. In figure 4.6, each edge's time windows (edge key) are presented in a table. Since there are three sensors, each table has three columns where each column represents the time window for each pair of sensors (the pairs being A-B A-C and B-C).

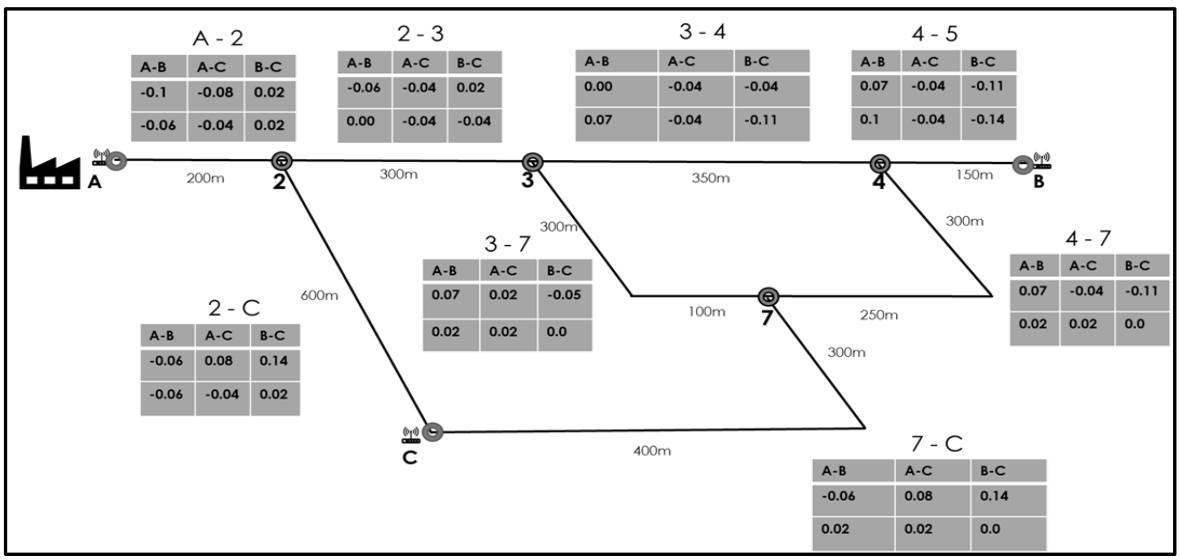

Figure 4.6 Example of distribution grid with three sensors

To determine the location of a fault, a *fault key* is compared to each edge's key (edge key) to find a match. A fault key is the delay between the fault's incident wave arrival time to each sensor/device. A match occurs when each component of the fault key falls within the corresponding time-windows of an edge key. An example of a fault key using figure 4.6 is [-0.05, -0.4, 0.01]. This fault key would match edge (2-3). This is because every component of the fault key falls within the upper and lower bounds for the corresponding components of the edge key for that segment.

After identifying the faulty segment of the cable, the exact location of that fault can then be determined. However, there are cases where either a single faulty segment cannot be identified, or the final fault location of the faulty segment cannot be determined because the placement and number of sensors was not enough to monitor the entire grid sufficiently. To acknowledge these cases, we introduce two types of localization, primary and secondary localization. Primary localization is where a specific point within a segment can be determined as the location of the fault. Secondary localization occurs when one or more segments are identified as potential locations for the fault, but it is not possible to pinpoint the fault's location. An algorithm for comparing a fault key to the edge keys to identify possible fault location segment (or segments) is provided in algorithm 1. Algorithm 2 determines if primary or secondary localization can be performed (using the results of algorithm 1) Algorithm 2 then carries out the calculation of the final fault location when primary localization can be performed.

Algorithm 1: Edge Match

---

1. **Input:** D, W //D is the fault key and W holds the time windows T for each edge
2. **Output:** M // a set to store all edges that match the fault key
3. **for each** e $\in$ E **do** //check each edge to look for matches to the fault key
4.   **for each** delay(i,j) $\in$ D //compare each time window and fault key component
5.   **if** ($T^{l}_{i,j}$ (e) $\le$ delay(i,j) $\le T^{u}_{i,j}$ (e)) //check for a match
6.   M = M $\cup$ {e} //add the edge to the set M if it matches
7.   **endif**
8.   **endfor**
9. **endfor**
10. **return M** //return all matching edges

Algorithm 1 describes the identification of one or more edges that could be where the fault is located. The input to Algorithm 1 (line 1) is the set of delays for each sensor pair for the fault signal, represented by *D* and the set of time windows for each sensor pair, *W*. For each edge, the value of *delay(i,j)* for each sensor pair (the fault key) is compared to time windows associated with sensors *i* and *j* (the edge key) (line 5) to determine if there is a match. A match occurs if *delay(i,j)* is within the range represented by the lower and upper bounds of each of the time windows as discussed in example 4 using figure 4.6. Edges in M represent the segments whose edge key matched the fault key.

Algorithm 2 is used to check if segmentation has allowed a single location to be determined for the fault location (primary localization) or if multiple locations were determined (secondary localization). With regard to primary and secondary localization, three cases need to be considered.

Case 1: $|M| = 1$ and $T^{l}_{i,j}(e) \neq T^{u}_{i,j}(e)$. This means that there is exactly one edge that matches the fault key with a time window where upper and lower bounds are not equal. The exact location of the fault is then calculated using equation 1. This case is classified as *primary localization* where an exact location has been determined.

Case 2: $|M| = 1$ and $T^{l}_{i,j}(e) = T^{u}_{i,j}(e)$: The edge has been identified but the upper and lower bounds of the time window that identifies that segment are equal. This means that the distance cannot be calculated, and an edge is highlighted as the location. This is classified as *secondary localization.* This case occurs when all points on the segment are calculated to be the same distance from the sensors *i* and *j* for the time window that identifies that segment. In this particular case there is only one segment that needs to be investigated to find the fault.

Case 3: $|M| > 1$: In this case there is more than one segment for which a pair of sensors have delay values that fall between the lower and upper bounds for those segments. This scenario is also classified as *secondary localization* since multiple edges are highlighted rather than a specific location. This case occurs if there are multiple segments where points on the segment are the same distance from the sensors *i* and *j*. This does not have to be all points but even a subset causes overlap of time windows.

For algorithm 2, the input is the values of *delay(i,j)* stored in *delay*, the distances between each pair of sensors stored in the variable *dist*, the fault signal speed *s* and the segments whose edge keys match the fault key generated by algorithm 1 *M*. Secondary localization is assumed to be the default (line 6). The set L is initialized to all edges in the set M (line 7). If M has more than one edge, then L is returned (line 18) as secondary localization occurred. If M only has one edge (line 8) the algorithm extracts that single edge (line 9) and checks if that edge has a time window where the lower and upper bounds are different (line 10). If this is the case, then equation 1 is used to calculate the location of the fault on that edge as a distance from the sensor i (line 11). The localization type is then changed to primary localization (line 12). The location of the fault and localization type (which is primary in this case) are then returned (line 13). Otherwise, L is returned and the *localizationType* is secondary localization (line 12). If M has multiple edges, then L is returned and the localization type is secondary (line 15).

Algorithm 2: Primary or Secondary Localization

1. **Input:** delay, s, dist, M //delay is the fault key, s is the propagation delay
2. //dist holds the distances between each pair of sensors, M is a set of edges
3. **Output:** L, localizationType, D
4. //L stores the edge or edges matching the fault key
5. // D stores the distance to the fault
6. localizationType = secondary
7. L = M; D=NULL
8. **if** (size(M) = 1)
9.   e = extract(M) //extract the only edge from the set M
10.   **if** ($T_{i,j}^{l}$ (e) $\neq$ $T_{i,j}^{u}$ (e)) //if a valid time window exists the distance to the fault is calculated
11.   D $= \mathrm{dist}(\mathrm{i,j})/2 + (\mathrm{delay}(\mathrm{i,j}) * \mathrm{s})/2$ //D is the distance to the fault from sensor i
12.   localizationType = primary
13.   return D, localizationType //The distance to the fault and primary localization is returned
14.   **else**
15.   **return** L, localizationType //L is an edge as the exact location was not found
16.   **endif**
17. **else**
18.   **return** L, localizationType //L is a set of edges as a unique edge was not identified
19. **endif**

Classifying edges as primary or secondary localization allows the identification of potential parts of the grid where the fault occurred.

4.5 Performance Evaluation

This section presents an evaluation of our approach. MATLAB was used to develop an electrical grid simulator using Simulink. A full electrical grid was simulated using Simulink electrical power components. In all tests a power generation component was used to simulate the generation of power of 120Kv at 60 hertz which was then placed on the transmission lines. A substation component was then used to step down the 120Kv power to 25Kv typically found in distribution grids. Finally, several transformers and loads were added to the distribution area of the model to represent the final end-user power draw on the grid. Sensors are then added as required by the test cases and test configurations described below.

4.5.1 Test Configurations, Test Cases and Evaluation

The distribution grid used for testing has 229 nodes and 243 edges. This graph is based on a map of a distribution grid in an Ontario town [Industry-Partner 2018] that came with GPS locations of various points on the power lines. We used these points as nodes. The edges between these nodes were then used as segments. To evaluate the solution, 60 *configurations* of the test grid were used. Each configuration uses the same grid. The configurations differ in the number of sensors (between 2 and 50) and location of the sensors. These are randomly generated (by using a random integer generator to generate the node IDs). For each configuration we randomly generated a fault location for each of the 243 edges (again using a random integer generator to choose how many meters along the edge a fault would be placed) which resulted in a corresponding fault signal. We then used our approach to localize the fault (243 times). These faults were the most common single phase faults which account for over 70% of all faults in distribution grids and are more challenging to localize as determined in [Sapountzoglou et al. 2019]. The faults were typical low impedance faults between 0.5 and 50 Ohms, which is within the range of low impedance faults used in literature [Sapountzoglou et al. 2019].

We then applied the following three tests on each configuration to evaluate the performance of the localization technique and edge classification: (1) Classify all edges as primary location edge (PLE) or secondary location edge (SLE) for each configuration and determine if the PLE and SLE classification was correct; (2) Determine if the correct segment is identified as the faulty segment for PLE faults and determine the accuracy of the PLE faults on that segment to the meter; (3) Determine the accuracy of SLE faults (if the set of edges returned contains the fault edge).

4.5.2 Performance

Table 4.1 shows a sample of 10 configurations taken from the 60 configurations used. For each configuration all edges in the graph were classified as PLE or SLE and then faults were placed on each edge at a random location as explained in the evaluation strategy. The fault was then localized to determine if it was on an edge classified as PLE or SLE. For all 60 configurations, a total 14,580 edges were classified with 60 different sets of sensor locations with a random number of sensors (between 2 and 50), and all PLE and SLE classifications were correct.

Table 4.1 shows a subset of the tests from multiple configurations of the accuracy of the classification of the PLEs and SLEs. To test the accuracy of PLE localization, 200 of the 14,580 edges classified as PLE across the 60 grid configurations were then used to determine the exact fault location using our fault localization approach. The results shown in table 4.2 are a sample of the 200 PLE accuracy tests. The error in meters was calculated as the difference between the actual location and the calculated location. The accuracy percent was calculated as the error divided by the length of the edge on which the fault was placed. The error on PLEs was a fraction of a meter no larger than 50 centimeters. This held true for all 200 tests generating an average over 99 percent. The error was investigated by reviewing the exact times recorded during tests and was found to be caused by rounding errors on the fractions of seconds used in the calculations. When identifying the faulty segment for the PLEs the correct segment was identified for every test resulting in 100 percent accuracy for locating the faulty segment of the grid.

Table. 4.1: Result samples from PLE and SLE classification tests.

Configuration	Number of Sensors in Test	Number of Classified PLE Edges	Actual PLE Edges	Number of Classified SLE Edges	Actual SLE Edges	PLE/SLE Classification Accuracy
1	29	23	23	220	220	100%
2	23	56	56	187	187	100%
3	24	41	41	202	202	100%
4	17	40	40	203	203	100%
5	31	35	35	208	208	100%
6	7	30	30	213	213	100%
7	20	59	59	184	184	100%
8	31	41	41	202	202	100%
9	9	34	34	209	209	100%
10	17	20	20	223	223	100%

Table. 4.2: Sample of localization accuracy for PLEs.

Test #	PLE Error (in meters)	Length of Fault Edge	PLE Location Accuracy
1	0.0097	561.80	99.999%
2	0.0122	396.70	99.996%
3	0.0075	122.60	99.992%
4	0.0198	87.30	99.977%
5	0.024	66.70	99.964%
6	0.0204	67.90	99.970%
7	0.3119	51.00	99.388%
8	0.0151	41.00	99.970%
9	0.0105	23.50	99.955%
10	0.0195	18.60	99.895%

SLE localization is the simpler of the two as it only requires the selection of the group of edges of which one edge contains the fault. The accuracy is based on whether the group of edges selected correctly encompassed the fault edge. Again 200 of the 14,580 edges classified as SLE edges were used across the 60 configurations. In all 200 SLE tests the algorithm correctly returned the fault edge within the group of edges returned generating an accuracy of 100%.

Overall, both the primary and secondary edge classification tests and fault localization tests showed each method to be highly effective. The solution presented allows highly accurate localization on all PLEs. Focus is now placed on when there are not an adequate number of sensors for full coverage of the grid, which results in SLEs (multiple fault locations). The solution presented by [Jamei et al. 2019] also focuses on not having an adequate number of sensors. Without an adequate number of sensors multiple fault locations are returned. With their solution, [Jamei et al. 2019] reduces the number of fault locations returned more effectively than previous methods in this scenario. The solution presented in this paper also addresses the scenario of inadequate monitoring and builds on this by allowing the user to know which areas of the grid will return multiple locations if a fault were to occur in those areas.

4.5.3 Test Model Parameters and Limitations

The structure of the model used in this work, as mentioned previously, is provided by an existing power company [Industry-Partner 2018]. Electrical grid models require a few basic components. The first is a power source which generates the power used by the consumers connected to the electrical grid. This power is generated at 60 Hz in North America and is distributed using three phases. Transformers are used to convert the power for transmission, distribution and for end user consumption. Using the structure provided by [Industry-Partner 2018], the gid's voltage, loads and transient signal propagation speed were set as listed below.

- Power Generating Station - 120 kV at 60Hz
- Stepdown Transformers - 120 kV to 25 kV
- Loads – 10,000 to 100,000W

- Transient Signal Propagation speed – 288,180,000m/s

The voltages, load and propagation speeds listed are not those of our industry partner but were set to standard values found in general distribution grids as that information (like the structure of the grid) is considered sensitive. The power source is a three-phase power generating station which generates 120,000 volts. This is then stepped down to a distribution grid voltage of 25,000 volts and transmitted through the structure provided by [Industry-Partner 2018]. The cables used for the grid were given a propagation speed of 288.18 million meters per second. The loads were set at values between 10 and 100 kilowatts to simulate consumer usage present in real world grids [Industry-Partner 2018]. The grid provided by [Industry-Partner 2018] also contains loops and branching elements making it sufficiently complex. Unfortunately we are unable to show the electrical grid provided by [Industry-Partner 2018] as it is not information that can be made public. However, we have constructed an example grid that contains the loops and branches that would usually be found in such a distribution grid. This is shown in Figure 4.7.

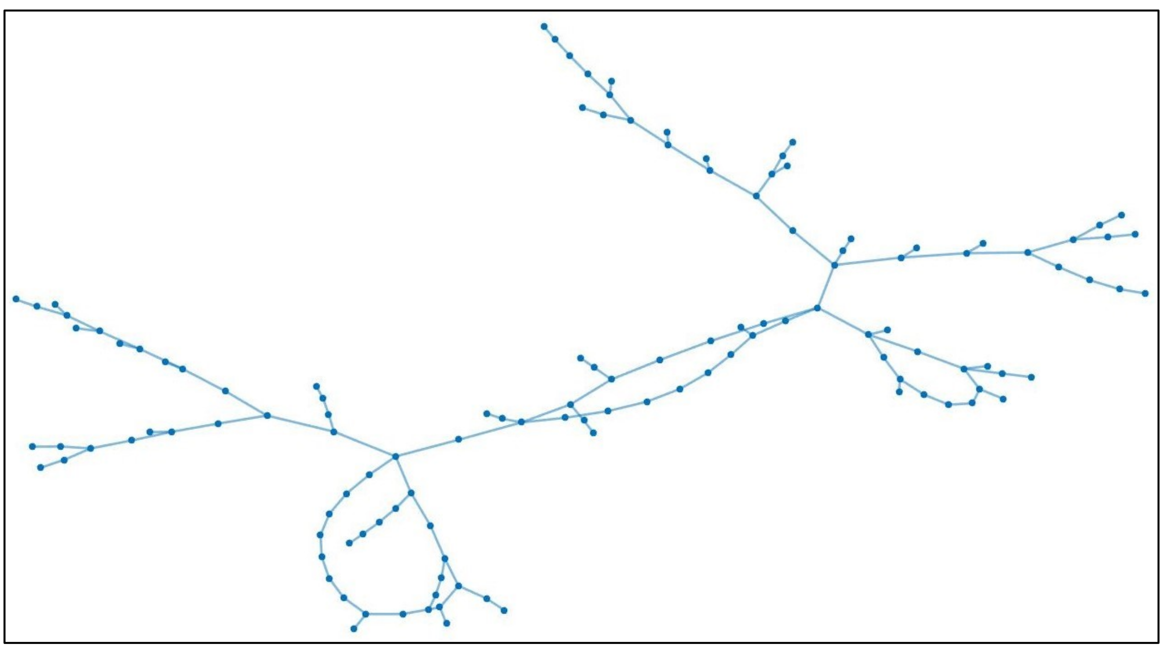

Figure 4.7 Example of an electrical grid with a complex realistic structure.

It must be noted a major limitation of work done in this area of research is in gaining access to real world grid models. There are IEEE test grids available but these grids were not designed to test fault localization techniques or even for analyzing travelling waves These test feeders were designed for power flow analysis [Schneider et al. 2017] . The significance of being designed for power flow analysis is that a power flow analysis does not require the distance between components to be considered with respect to the propagation time of transient signals or in most cases does not require the consideration of transient signals at all. Using test grids that are known to be an inaccurate representation of a simulation of transient signals may invalidate the success of the testing performed. We therefore refrained from using these test feeders. In the absence of accurate models there is a lack of test cases to use for transient analysis and fault localization method development. The lack of real-world models of test grids for transient analysis limits the testing of fault localization solutions to those attained from an industry partner which are not provided for public use. This research utilized such a source but acknowledges a need for a set of publicly available test grids for fault localization method testing and development as this would allow clear comparisons to be made between different methods.

4.5.4 Comparison with Related Work

Comparisons between different fault localization methods are difficult due to the lack of a standard set of public distribution grids. It is possible that a solution may perform better on one test grid due to the topology or attributes of that grid (like average load and number of customers). However, it is important to provide an evaluation with existing work. A comparison can be provided with solutions that provide a similar evaluation method even though they do not share the same test grid. In our evaluation we perform two measurements that can be used for a comparison:

- How well the solution identifies the correct segment for a fault.
- How accurate the actual location of the fault is calculated on the identified segment

The solution presented in [Sapountzoglou et al. 2019] performs a similar evaluation of their fault localization solution. The solution in [Sapountzoglou et al. 2019] was highlighted in Chapter 2 as a very promising approach to localizing faults in distribution grids. The solution in [Sapountzoglou et al. 2019] is evaluated on its ability to accurately identify a faulty sector and how well the distance to the fault is estimated (calculated as a percentage error). With this common evaluation, we compare the solution presented in this chapter

(Chapter 4) with the solution presented in [Sapountzoglou et al. 2019]. For this evaluation we only consider primary localization and therefore assume that there is an adequate number of devices/sensors in the grid.

On the grid used by [Sapountzoglou et al. 2019], the average accuracy for single phase faults (less than 50 Ohms) on the fault distance was 93 percent for the solution developed by [Sapountzoglou et al. 2019]. The average fault accuracy for our solution was 99 percent. For the identification of a faulty segment, Sapountzoglou's solution [Sapountzoglou et al. 2019] had an average of 81 percent where our solution had 100 percent accuracy for faulty segment identification. These results show that the approach taken in this chapter is more effective at identifying faulty segments. It is also more accurate when calculating the fault distances. However, there are some limitations in what can be deduced from this comparison. For Sapountzoglou's solution [Sapountzoglou et al. 2019] the faults were under 50 Ohms but the exact values may have a significant impact on performance. This is hinted to be the case by [Sapountzoglou et al. 2019] as the faults with impedance greater than 50 Ohms were significantly harder to localize. There are also differences in the complexity of the grid as Sapountzoglou's test grid [Sapountzoglou et al. 2019] has significantly less branches which may impact a solutions performance.

What can be taken from this comparison is that the solution presented in this chapter (Chapter 4) has significant potential to outperform other fault localization methods. The next test should be done with both solutions implemented on the same electrical grid. The tests should also use the same simulation software and the comparison should be made between individual test instances rather than an average.

4.6 Conclusion and Future Work

Our work provides an approach to fault localization in a distribution grid based on multiple sensors, segmentation and the use of delays representing when sensors receive the incident wave caused by a fault. This enables us to localize the fault when there are many possible locations for the fault. Our evaluations show that the approach has huge potential based on results that show that the fault localization solution can very accurately localize faults on a distribution grid.

Our results suggest that the total number of PLEs is not correlated with the number of sensors used. In the results shown in table 4.1, the total number of PLEs in in each configuration did not appear to be correlated with the number of sensors used. For example, taking configurations 1 and 2 as examples, this is apparent. In configuration 1 it is seen that 29 sensors only resulted in 23 PLEs. On the other hand, configuration 7 only used 20 sensors but generated 59 PLEs. The placement of sensors was random (using a random integer generator to generate the node IDs). This suggests that the effectiveness of our approach depends on the location of the sensors in the grid.

Generally, for fault localization to occur the grid must be energized. A few seconds after a fault occurs, the grid is powered down and is no longer active. It is possible that there can be multiple areas of a distribution power line being damaged. However, due to the speed of the fault signals, it is improbable that the faults causing the damaged areas would occur while the grid is still energized. Therefore, in scenarios with multiple damaged areas, solutions that localize faults like the one presented in this paper, do so for the first fault which is the fault that causes the power outage to occur. For future work there are two main areas to explore. The first area is determining the optimal locations for the devices to maximise the performance of the localization method. The second area is in assisting the optimization process in SLEs which would be common when devices are limited.

With respect to determining the sensor locations, the ability to differentiate between areas of the grid that would allow precise fault localization is of particular importance. The ability to classify these different areas allows decisions to be made on where sensors should be placed to maximize the solution’s performance. Focus will be placed on developing algorithms that allow for the optimal placement of sensors that maximize coverage under the primary localization scheme (maximizing the number of PLEs). A higher number of PLEs means that fewer areas have to be searched for finding the fault.

The application of historical and equipment lifespan data to secondary localization edges (SLEs) is a possible area for future work. Information about the age of components and the location of previous faults can be used when a set of locations are generated to create a prioritized list of the set of locations. Locations in the set that have previously had faults or have the oldest equipment could be identified as the most likely locations for the fault. To implement this method there is a need to first establish the existence of a link between previous faults or the age of equipment and the occurrence of faults.

4.7 Summary

The method of localizing faults presented in this chapter builds on the initial work done (in Chapter 3) on synchronizing travelling wave data when using multiple devices. The fault localization method presented in this chapter (Chapter 4) utilizes a graphical representation of the distribution grid to separate the grid into segments. Firstly, time windows are established to identify each segment of the grid. Secondly, a fault key is generated from the synchronized travelling wave data (made available using the contributions of Chapter 3). Finally, a comparison between the fault key and the time windows of each segment is performed. This comparison allows a single segment and fault location to be identified. This fault localization method is a significant contribution to research in fault localization. Previous fault localization solutions in distribution grids struggled to generate a single fault location due to the complex nature of the distribution grid. The solution presented in Chapter 4 addresses the limits of other work by generating a single fault location, given that an adequate number of devices are placed in the distribution grid.

In this chapter a classification scheme is also introduced to determine there is an adequate number of sensors to ensure a single fault location can be generated. This classification scheme is leveraged in Chapter 5. In Chapter 5 focus is now placed on optimizing the location and number of devices used to capture travelling waves, to maximize the performance of the fault localization method presented in this chapter (Chapter 4).

4.8 Bibliography

Andrade, L. d. and T. P. d. Leão. Impedance-based fault location analysis for transmission lines. *PES T&D 2012*, (2012).

Crossley, P. A. and P. G. McLaren. "Distance Protection Based on Travelling Waves." *IEEE Transactions on Power Apparatus and Systems* **PAS-102**(9): 2971-2983, (1983).

Das, S., S. Kulkarni, N. Karnik and S. Santoso. Distribution fault location using short-circuit fault current profile approach. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Gabr, M. A., D. K. Ibrahim, E. S. Ahmed and M. I. Gilany. "A new impedance-based fault location scheme for overhead unbalanced radial distribution networks." *Electric Power Systems Research* **142**: 153-162, (2017).

Girgis, A. A., C. M. Fallon and D. L. Lubkeman. "A fault location technique for rural distribution feeders." *IEEE Transactions on Industry Applications* **29**(6): 1170-1175, (1993).

Industry-Partner. Ontario Electricity Provider, (2018).

Jamei, M., R. Ramakrishna, T. Tesfay, R. Gentz, C. Roberts, A. Scaglione and S. Peisert. "Phasor measurement units optimal placement and performance limits for fault localization." *IEEE Journal on Selected Areas in Communications* **38**(1): 180-192, (2019).

Jun, Z., D. L. Lubkeman and A. A. Girgis. "Automated fault location and diagnosis on electric power distribution feeders." *IEEE Transactions on Power Delivery* **12**(2): 801-809, (1997).

Karnik, N., S. Das, S. Kulkarni and S. Santoso. Effect of load current on fault location estimates of impedance-based methods. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Magnago, F. H. and A. Abur. "Fault location using wavelets." *IEEE Transactions on Power Delivery* **13**(4): 1475-1480, (1998).

Marguet, R. "Improved fault localization method for electrical power distribution networks." (2015).

Mwifunyi, R. J., M. M. Kissaka and N. H. Mvungi. "Distributed approach in fault localisation and service restoration: State-of-the-Art and future direction." *Cogent Engineering* **6**(1): 1628424, (2019).

Pourahmadi-Nakhli, M. and A. A. Safavi. "Path Characteristic Frequency-Based Fault Locating in Radial Distribution Systems Using Wavelets and Neural Networks." *IEEE Transactions on Power Delivery* **26**(2): 772-781, (2011).

Qin, J., X. Chen and J. Zheng. Travelling wave fault location of transmission line using wavelet transform. *Power System Technology, 1998. Proceedings. POWERCON '98. 1998 International Conference on*, (1998).

Saha, M., F. Provoost and E. Rosolowski. "Fault location method for MV cable network." (2001).

Sapountzoglou, N., B. Raison and N. Silva. Fault Detection and Localization in LV Smart Grids. *2019 IEEE Milan PowerTech*, (2019).

Schneider, K. P., B. Mather, B. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa and L. R. de Araujo. "Analytic considerations and design basis for the IEEE distribution test feeders." *IEEE Transactions on power systems* **33**(3): 3181-3188, (2017).

Seung-Jae, L., C. Myeon-Song, K. Sang-Hee, J. Bo-Gun, L. Duck-Su, A. Bok-Shin, Y. Nam-Seon, K. Ho-Yong and W. Sang-Bong. "An intelligent and efficient fault location and diagnosis scheme for radial distribution systems." *IEEE Transactions on Power Delivery* **19**(2): 524-532, (2004).

Srinivasan, D., R. L. Cheu, Y. P. Poh, A. Kim and C. Ng. "Automated fault detection in power distribution networks using a hybrid fuzzy±genetic algorithm approach." (2000).

Wen, F. and C. S. Chang. "A new approach to fault diagnosis in electrical distribution networks using a genetic algorithm." *Artificial Intelligence in Engineering* **12**: 69-80, (1998).

Chapter 5

5 Optimization of Sensor Deployment in Electrical Distribution Grids for Fault Localization

5.1 Introduction

There are many fault localization strategies that have been developed for both the transmission and distribution grid. The majority of these can be classified as either impedance-based (e.g., [Saha et al. 2001],[Das et al. 2011],[Karnik et al. 2011],[Takagi et al. 1982]) or travelling waves (e.g., [Crossley and McLaren 1983],[Magnago and Abur 1998],[Hizam and Crossley 2006]) Some of the more promising solutions extend these initial solutions by employing multiple sensors [Trindade et al. 2014],[Trindade and Freitas 2017],[Škumát and Ž 2019],[Sapountzoglou et al. 2019]. The method presented in [Hunte et al. 2021] follows this successful trend and employs multiple sensors as presented in Chapter 4. A description of the solution is provided in the following paragraphs to provide context for the optimization problem. The solution takes a travelling wave-based approach. It reduces the complexity of the signal analysis done on the travelling waves through multi-device/sensor synchronization and graph segmentation using time windows.
The solution can be summarized as follows:

- Segmenting the electrical grid
- Establishing time-windows for each segment
- Upon the occurrence of a fault identify the faulty segment using the time windows
- Calculate final fault location on the faulty segment

To determine the location of a fault, the delay between that fault's arrival time to each sensor/device (referred to as the fault key) is compared to the time windows of each edge (referred as the edge key) to find a match. A match occurs when the fault key falls within the time-windows of an edge key. The match would indicate which edge of the cable the fault occurred.

After identifying the faulty segment of the cable, the exact location of that fault can then be determined. However, there are cases where either a single faulty segment cannot be identified or the final fault location on the faulty segment cannot be determined because the placement and number of sensors was not enough to cover the entire grid sufficiently. To acknowledge these cases, we introduced two types of localization, primary and secondary localization.

Primary localization occurs when only one segment matches the fault key and one of the time-windows of that segment is valid. When one segment matches the fault key, that segment can be identified as the fault segment. After this, a valid time window is needed to determine the exact fault location on the fault segment. A valid time window is a time window that has a different lower and upper bound. If more than one segment matches the fault key or there is no valid time window, only a fault area can be determined. A fault area is the segment or segments that match the fault key. This is referred to as secondary localization.

There are three cases that may occur during the localization process:

Case 1 (Primary Localization) – A single segment matches the fault key with a valid time-window.

Case 2 (Secondary Localization) – A single segment matches the fault key and there is no valid time window.

Case 3 (Secondary Localization) – Multiple segments match the fault key.

Using these three cases, each segment can be classified as a primary or secondary localization segment. Each segment can be classified by determining if one of a segment's time windows is unique and if there is a valid time window that has a different upper and lower bound).

Segment classification is dependent on the number and location of the sensors in the grid (referred to as a configuration). The link between segment classification and a given configuration of the grid provides a way of evaluating the performance of a configuration. The classification of the segments can be used to determine which parts of the electrical grid are covered by either primary or secondary localization. It ultimately shows how well faults can be localized with a given number and location of sensors.

Primary segment classification can establish a metric that correlates to the effectiveness of the placement of the sensors. The amount of power cable in the grid (in meters) that is classified under primary localization segments is used as the performance metric. This metric can then be used to compare the performance of different sensor quantities and locations. We use this to optimize the number and location of sensors, reducing the cost of localizing faults.

5.2 The Problem Description

In this section the optimization of sensor placement problem is defined as a subset problem [Qian et al. 2017]. The size of the problem's search space is also shown to provide perspective of the work to be done to find an optimal solution. In addition to this, a definition of the objective function used to determine a solution's performance (towards the optimization goal) is also provided. This function is used by both optimization methods explored to provide a consistent metric for the performance of solutions generated by the optimization methods.

5.2.1 The Objective Function

Given an undirected graph $G = (V, E)$, each vertex/node $v \in V$ is a possible location for a sensor in the grid and each edge $e \in E$ is the electrical cable between each adjacent node (possible sensor location).

For clarification on what the graph G represents, refer to figure 5.1 which shows the example grid originally shown in section 4.5. Each node in figure 5.1 is numbered (not shown) and is an element in the set V and the edges E are the paths (electrical cable) that exist between those nodes.

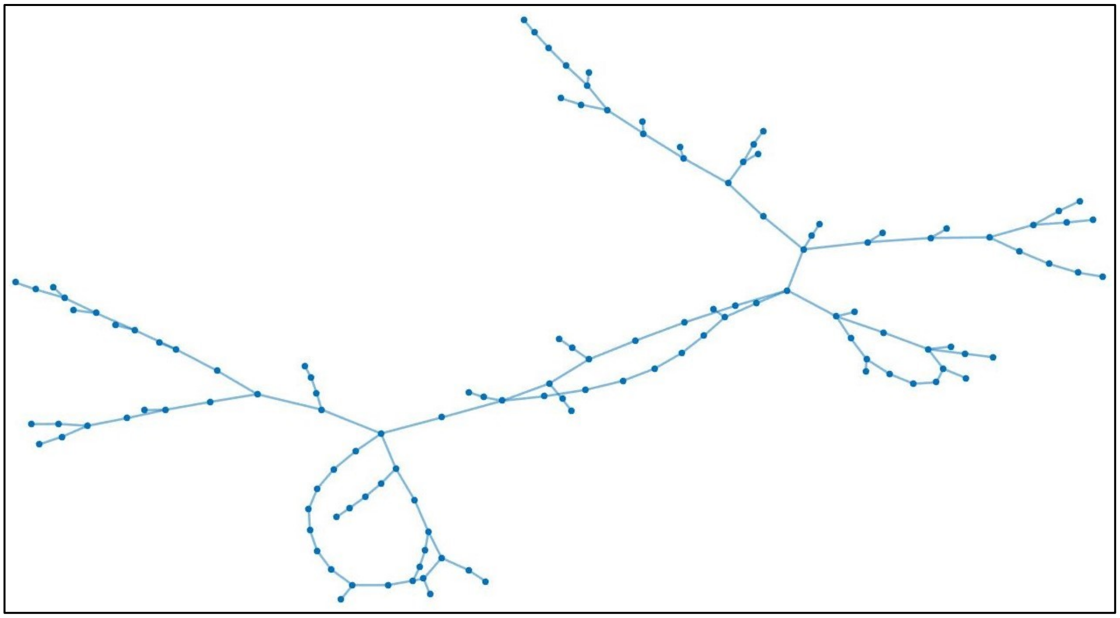

Figure 5.1 Example of an electrical grid with a complex structure

Let the set T contain the time windows (given as the upper bound $T_{i,j}^{u}$ and lower bound $T_{i,j}^{l}$) for all pairs i, j of V, for each edge $e \in$ E. An edge $e \in$ E qualifies as a primary localization edge if it meets the following criteria:

- When the edge e is compared to each other edge there exists at least one unique time window ($T_{i,j}^{l}(e)$, $T_{i,j}^{u}(e)$) in every comparison that does not overlap (the edge is uniquely identifiable).
- For that edge e there is a time window where the upper and lower bounds are not the same (there is a valid time window).

The primary localization criteria can be formulated as follows:

- $\exists\ T_{i,j}^{l}(e) \in T$(lower bound) and $T_{i,j}^{u}(e) \in T$(upper bound) such that $\forall\ k \in E, k \neq e$ and $\nexists\ \mathrm{T}_{i,j}^{l}(e) \leq \mathrm{T}_{i,j}^{l}(k) \leq \mathrm{T}_{i,j}^{u}(e)$ and $\mathrm{T}_{i,j}^{l}(e) \leq \mathrm{T}_{i,j}^{u}(k) \leq \mathrm{T}_{i,j}^{u}(e)$
- $\mathrm{T}_{i,j}^{l}(e) \neq \mathrm{T}_{i,j}^{u}(e)$.

The optimal sensor placement problem can be formulated as an optimization problem in which one aims to maximize (or minimize) a user-defined objective function related to the characteristics of a structural system, where the sensor locations are defined as the discrete optimization variables (parameters) subject to a constraint. It is important to note that the constraint used by the optimization problem may be redefined by industry. As the constraint is unknown, a reasonable assumption is made to define the constraint which is that there will be a maximum number of sensors to be placed in the grid. For the testing performed in this chapter the constraint used is a maximum number of sensors. However, industry may have a different view or views on what acceptable solution may be. To account for this variability, the formulation of the optimization problem is kept as general as possible.

In this work, *sensor placement* is a finite subset of locations S from V where V represents all possible locations for placing sensors in the distribution grid. The goal is to maximize the amount of cable classified as a primary localization edge (PLE). The problem can be formulated as follows:

$$\max \sum_{i=1}^{n} f(S, i) * w(i)$$

$subject\ to$

$$\mathrm{Cost(S)} < \mathrm{B}$$

Where *n* is the number of edges and $f(S, i)$ returns 1 if the edge *i* is a PLE and 0 if it is not. *w(i)* returns the weight of the edge *i*. B represents the budget, and the cost function Cost(S) evaluates the cost of the sensors used in relation to the budget B. Essentially, the problem is to select a subset, S,of V that maximizes the objective function. The constraint on the choice of S is that the cost of the selection made (given by cost of S) be less than some threshold (budget) B. Multiplying the return value of f by the weight of the edge allows the total amount of cable that satisfies the primary localization criteria to be calculated. This weight allows not just the length of the cables to be considered but can also allow priorities (in monitoring certain areas of the grid) to be incorporated if needed.

This is a finite combinatorial optimization problem, so one way to solve it is by brute force. This requires enumerating all possible subsets where cost(S) is within the budget, evaluating *f*, and picking the best subset. The number of possible subsets is very large. The number of possible subsets grows extremely fast as V increases, so the brute force approach quickly becomes infeasible as V becomes large. The problem in our work is considered to be a subset problem which is NP complete [Hamo and Markovitch 2005].

5.3 Related Work

The work in this chapter presents an optimization algorithm for the specific problem of determining the most effective locations for the sensors used to localize faults for the fault localization solution in [Hunte et al. 2021]. As the problem does not scale well, a simple solution was developed using a greedy approach to choosing locations. This solution was then evaluated against a more sophisticated and computationally expensive approach, namely a genetic algorithm. Both methods have been studied as possible solutions to finding optimal or near optimal solutions. We focus on these two methods because of the need to look at the resources used when generating solutions. The greedy approach provides a computationally inexpensive approach and the genetic algorithm provides a more expensive approach to solving the optimization problem.

5.3.1 Greedy Algorithms

Greedy algorithms often only find locally optimal solutions, but can provide decent approximations to problems [Cormen et al. 2009]. A detailed explanation on greedy algorithms is provided by [Cormen et al. 2009] who argue that in many cases greedy algorithms are able to find the globally optimal solution. There are several types of greedy algorithms stated by [Lin et al. 2013] who also note four of the most commonly used types of greedy algorithms are:

- Pure greedy
- Relax greedy
- Orthogonal
- Stepwise projection

An in depth technical look at these main types of greedy algorithms are provided in [Dereventsov 2012] [DeVore and Temlyakov 1996] [Temlyakov 2003] [Barron et al.

2008]. The efficiency of these methods are discussed by [Temlyakov 2003] as it relates to approximation and outlines some variations of these methods.

5.3.2 Genetic Algorithms

Optimization techniques are categorized as being calculus-based, enumerative or guided random [Bandyopadhyay and Saha 2013]. We pay particular attention to guided random techniques and specifically genetic algorithms due to their ability to handle problems where the search space is large and near optimal solutions are acceptable, which is the case for our optimization problem. We can use genetic algorithms as the placement of sensors is a combinatorial optimization problem. A genetic algorithm (GA) is an iterative, reinforcement learning, guided search technique that explores a search space, based on survival of the fittest to find optimal or near optimal solutions to combinatorial optimization problems. A concise survey of GAs is conducted by [Srinivas and Patnaik 1994]. A large source of information on GAs is provided in [Ghosh and Dehuri 2004] that shows the considerations one has to make when employing GAs.

GAs are a very effective means of finding optimal or near-optimal solutions to combinatorial optimization problems [Bandyopadhyay and Saha 2013]. They intelligently navigate through the possible solutions of a given problem in search of an optimal solution [Bandyopadhyay and Saha 2013]. GAs encode an optimization solution into a chromosome. A chromosome is a representation of a solution in a form (like a string of integers) that allows it to be easily manipulated by the following genetic operations:

- Initialization - Creating an initial random set of solutions(chromosomes) called a population.
- Fitness Evaluation (performance)– Determines how well a solution performs against the defined objectives.
- Selection – This refers to solutions (chromosomes) used to generate the new population (based on fitness).
- Crossover – This is the process by mixing parts (genes) of chromosomes to generate a new chromosome.
- Mutation – This introduces random parts (genes) to maintain a diverse gene pool.

The following algorithm shows how these operations are used in a genetic algorithm during its execution.

Algorithm 1: Basic GA

```
1.    Create E (initially nil) for elite performers
2.    Initialize population set P
3.    While termination criterion not met do
4.        Evaluate(P)
5.        Update E with best performers in P and E
6.        Set X = Selected solutions from E
7.        Set P = Crossover and Mutated set X
8.    Endwhile
```

Line 1 creates an empty set for elite performers. The elite performers are the best performing solutions found by the algorithm. Line 2 initializes the population with randomly generated solutions which are usually random integers. The rest of the algorithm (lines 4 to 7) executes a loop until the criteria for termination (e.g., number of iterations; a certain fitness threshold) is met. In this loop the population (set of solutions) is evaluated to determine how well they have performed in relation to the optimization goal (line 4). E is updated to hold the best performing solutions (line 5). The elite solutions are then selected to generate a new population via the crossover operation (line 6). A new generation of solutions created by the crossover and mutation operations replaces the solutions in P (line 7). Once the loop ends the best performing solutions found during the algorithm's execution would be found in E. The best performing solution in E is selected as the final solution.

The greedy based approach is evaluated in [Vafaie and Imam 1994] against a genetic algorithm approach for feature selection (determining which features should be used from a set of data for making decisions about that data). They found that when it comes to feature selection, greedy-like searches get trapped in local peaks but can be more efficient in some cases. The work presented in [Vafaie and Imam 1994] also found that the GA-based method was able to improve the robustness of the feature selection at the expense of increased computational complexity. The work in this chapter further explores the performance of greedy and genetic algorithms in a different problem. We specifically compare the two in optimizing sensor locations for the solution developed in [Hunte et al.

2021] with the end goal being the selection of a cost-efficient optimization algorithm to reduce the resources needed by the fault localization solution.

5.4 The Evaluation Function

To find optimal or near optimal solutions the need to compare solutions is necessary. We now present an evaluation function that allows a score to be provided for a given solution. This function can be used by both the greedy and genetic algorithms to compare the quality of different solutions. For this evaluation function, a possible solution to the optimization problem is given as a set S of nodes (sensor locations) where $S \subseteq V$ and V is the set of all nodes in the electrical grid's graph. The set S is of size n, with n representing the number of sensors to be placed in the grid. The set S is used to generate the set *TW* of time windows for each pair of sensors for each edge as discussed in [Hunte et al. 2021]. The total PLE coverage (sum of the cable lengths of the primary localization edges) is used as the score for the set S for the evaluation function and tests performed in this chapter. The sum of the cable lengths is used as the score for a solution because each edge may have a different length. If the lengths of cables are not considered, then an edge representing a 10km cable and an edge representing a 1km cable will be treated the same (each having a value of 1) by the optimization algorithm. The optimization algorithm would then be unable to distinguish between a solution where the 1km cable is covered under primary localization and the 10km cable is covered. In reality, the solution covering the 10km cable is the better option. To avoid this, the lengths of cables are considered in the score for the solutions by adding weights to each edge which is the length of cable for that edge in the electrical grid. A solution's score is then calculated by adding the *weights* of the PLEs in the graph.

Algorithm 2: Evaluate

```
1.  Input:  S, V, E //set of sensor locations S and vertices V and edges E
2.  Output:  total_PLE //total cable classified as primary localization
3.  numEdges = size ( E )
4.  TW[][] // 2-D array of edge objects to hold the set of time windows for each edge
5.  TW = generateTimeWindows( S, V, E ) //create a set of time windows for each edge
6.  Set total_PLE to 0 and edgeCount to 0
7.  for i = 0 to numEdges do //select an edge i to see if it is uniquely identifiable
8.    edgeCount = 0
9.    for j = 0 to numEdges do //iterate through all other edges j
10.      if (i != j) //do not compare the time windows for the same edge
11.         for m = 0 to size( timeWindow ) do //iterate through all time windows m
12.            //check to see if the time window does not have an overlap
13.            if (TW[j][m].lower > TW[i][m].upper) OR (TW[j][m].upper < TW[i][m].lower)
14.                 edgeCount = edgeCount + 1 // a unique time window has been found
15.                 break //no need to continue checking other time windows
16.              endif
17.            endif
18.            if ( TW[i][m].upper != TW[i][m].lower ) //check if there is a valid time window
19.              validTWExists = true //a valid time window exists for edge i
20.            endif
21.          endfor
22.      endif
23.   endfor
24.   if (edgeCount == numEdges) //the edge i is uniquely identifiable
25.      if(
26.      total_PLE = total_PLE + length(E[i])
27.   endif
28.  endfor
29. return total_PLE
```

Each edge represents a segment of a power line and is associated with a set of time windows for each pair of sensors. Algorithm 2 (the Evaluate function) takes as input a set S of sensor locations and the vertices and edges that represent the grid (line 1). The Evaluate function returns the total amount of cable classified as primary localization. The algorithm generates a set of time windows (*TW* in line 5). Drawing reference from Section 5.2.1, the variable *TW* holds the time windows ($T^{l}_{i,j}$ (e), $T^{u}_{i,j}$ (e)) for each pair of sensors *i, j* for each edge e $\in$

E. Here, the first dimension of *TW* is used to access the edge and the second dimension is used to access a single time window within that edge. The outer for loop (line 7) goes through each edge to check if edge *i* has a time window with no overlap. The second for loop starting on (line 9) allows the current edge *i* to be checked against all other edges. The third for loop iterates through each time window for the edges being compared (line 11) to look for a unique time window. Inside the third loop is an if statement (lines 13 to 16) which increments *edgeCounter* if a unique time window is found for edge i. Recall from Section 5.2.1 that an edge $e \in$ E qualifies as a primary localization edge if it meets the following criteria:

- When the edge *e* is compared to each other edge there exists at least one unique time window $(T_{i,j}^{l}(e),\ T_{i,j}^{u}(e))$ in every comparison that does not overlap (the edge is uniquely identifiable).
- For that edge *e* there is a time window where the upper and lower bounds are not the same (there is a valid time window).

If a unique time window is found for edge *i* for each other edge, then edge *i* satisfies the first criteria for primary localization (being uniquely identifiable). If the variable *validTWExists* is true, then the second and final criteria for primary localization (having a valid time window) is also true and the edge *i* would be a PLE. As such the length of edge *i* is added to *total_PLE*. At the end of the Evaluate function the variable *total_PLE* contains the total length of all PLEs. This evaluation function is used in Section 5.5 in a greedy algorithm and in Section 5.6 in a genetic algorithm to determine the performance of a given solution. In our evaluation function we use the weight of each edge to convert the number of PLEs to a total amount of cable to account for edges/segments being of different lengths rather than just a count of the number of PLEs.

5.5 A Greedy Algorithm for Optimization of Senor Locations

The greedy algorithm presented in this chapter takes an incremental approach to optimizing sensor locations (adding one sensor at a time). However, the requirement for the problem is that there must be no less than two sensors. The origin of this is in the fact that the localization technique in [Hunte et al. 2021] requires at least two sensors. Therefore, the placement of two sensors must be done first and then the general greedy algorithm can be executed to add additional sensors until the desired n sensors are placed. The greedy algorithm is shown in algorithm 3.

The greedy algorithm takes as input the number of sensors to be placed in the grid and the nodes in the graph (potential sensor locations) (line 1). It then calculates the locations for the initial two sensors by using an exhaustive search of all possible pairs of nodes to find the globally optimal locations for the first two sensors (maximizing the amount of cable classified as being PLEs) (line 4) and removes these first two nodes from the list of potential locations in Nodes. The algorithm then enters a series of loops. The inner loop takes the set of sensor nodes and searches through all remaining potential sensor locations in the nodes variable to determine the next sensor's best location (lines 7 to 14). The outer loop repeats the addition of a sensor location until the desired number of sensors have been placed (line 5).

Algorithm 3: Greedy

```
1.  Input:  numberOfSensors, Nodes
2.  Output:  locations
3.  locations = []
4.  [locations Nodes] = initialSensorPair(Nodes) //generate first 2 locations
5.  for j=1: numberOfSensors - 2  do //choose the remaining n-2 sensor locations
6.     bestPerformance = 0
7.     for i=1:numberOfLocations do //find the next best sensor location
8.        sensorNodes = [locations Nodes[i] ] //add the ith location for evaluation
9.           performance = Evaluate( sensorNodes ) //check performance of the ith location
10.          if ( performance > bestPerformance )//update the current best location if needed
11.                bestLocation = Nodes[i]
12.                bestPerformance = performance
13.          endif
14.       endfor
15.    locations = [locations bestLocation] //add the nest best location found to the list
16.    endfor
17.    return locations //return final set of sensor locations
```

5.6 A Genetic Algorithm for Optimization of Sensor Locations

The genetic algorithm used for the optimization of sensor locations for n sensors is outlined in Algorithm 4. It uses traditional genetic operations to navigate the search space and find optimal or near optimal solutions. An explanation of the operations performed by the genetic algorithm is provided in Section 5.6.1. An example to clarify the genes and chromosomes used by the genetic algorithm is given in Section 5.6.2. The definitions of the functions used in the genetic algorithm are provided in Section 5.6.3.

5.6.1 Genetic Algorithm Operations

Algorithm 4 shows the operations performed by the genetic algorithm. In this section an explanation of the algorithm's design is provided. In this explanation the term random is used to refer to the use of a random number generator. The precise use of random number generation is provided in Section 5.6.3 when the details of the functions used in the genetic algorithm are presented.

The genetic algorithm takes the number of sensors to be placed in the grid as input. Next E is initialized to hold the ten best performing solutions found (line 3). The size of the elite set was tested with values from 1 up to 100 (the size of the population) in increments of ten. An elite set size of 10 was chosen as it is the smallest size that consistently provided optimal or near optimal solutions.

On line 4, the variable *max* is set to fifty so that the while loop runs until there has been no change to E for fifty generations. This was chosen because the genetic algorithm uses a dynamic probability of mutation (*pm*) when convergence is detected. The probability of mutation is adjusted according to the rate of convergence similarly to the method presented in [Srinivas and Patnaik 1994]. Convergence occurs in this scenario when there is no change in the elite set. The genetic algorithm presented starts with a mutation rate of 0.02 and when convergence is detected the mutation rate is increased by 0.02 up to the maximum value of one. This would require fifty iterations to increase the mutation rate to its maximum value.

Algorithm 4: Genetic

```
1.  Input:  numberOfSensors, numberOfNodes, Nodes
2.  Output:  locations
3.  E = null //2D array to hold elite chromosomes
4.  convergeCounter = 0, max = 50, pSize = 100
5.  P = initialize( numberOfSensors , pSize ) //create initial population
6.  while  convergeCounter < max do //execute until the algorithm has converged
7.     for i = 1:size(P)  do //evaluate each chromosome (solution)
8.         performance(i) = Evaluate( P(i) ) //function as defined in section 5.4
9.     endfor
10.    prevE = E //temporarily store old elite set to check for change
11.    E = update(performance, E, P) //update E with best solutions in population
12.    if ( prevE = E ) //check for a change in elite set (detect convergence)
13.       convergeCounter = convergeCounter + 1
14.    else
15.       convergeCounter = 0 //reset counter whenever a change is detected
16.    endif
17.    X = select(E, P, numberOfSensors) //choose chromosomes for crossover
18.    P=crossoverAndMutate(X, convergeCounter, numberOfSensors, numberOfNodes, pSize)
19. endwhile
20. return locations = max( E ) //return best solution
```

The population size was set to 100 (line 4) as it is considered a reasonable small population size by [Srinivas and Patnaik 1994] and as such would use a smaller amount of resources than larger population sizes, maintaining a focus on the efficiency of the optimization process. The population P is therefore initialized with a set of 100 chromosomes (line 5).

The while loop (lines 6 to 19) allows:

- The solutions in P (the current generation) to be evaluated (line 7 to 9)
- The elite set E to be updated with the ten best solutions (line 11)
- The solutions in E and P to be selected for crossover for creating the next generation of solutions (line 17)

- The genetic operators (crossover and mutation) to be performed to create new generations (line 18)

Once the loop terminates the best solution in the elite set is selected as the final solution.

Selection is performed (on line 17) by choosing random chromosomes (by using a random integer generator to generate the indices of the chromosomes) from the elite set E and members of the current population P to create the set chromosomes needed for the crossover operation (creating the next generation). The crossover probability (*pc*) was set to one for each iteration. Setting *pc* to one ensures that all members of the next generation will be a result of the crossover operation. The maximum value of one was used for the crossover probability as it is a common value used when implementing genetic algorithms [Srinivas and Patnaik 1994].

The *crossoverAndMutate* function (line18) creates the next generation of solutions from the set returned by the *select* algorithm. The number of chromosomes to be mutated is controlled by the value of *pm*. The value of *pm* is used as a percentage of the population size to determine the number of new chromosomes that will be mutated. The initial value of *pm* would be 0.02 until convergence is detected. When convergence is detected (no change in the elite set) the initial value of 0.02 would be multiplied by convergeCounter (passed into the crossover and mutate function on line 18) to provide the value of *pm*.

If the new chromosome is to be mutated, each of the genes are chosen randomly (by using a random integer generator) from all possible values for a gene (any node) as long as it does not create a duplicate gene in the chromosome. If it is not a mutated chromosome then two random chromosomes (by using a random integer generator to generate the indices of the chromosomes) are chosen from X (as a pair of parents). A new chromosome is created from the genes of that pair unless it is selected for mutation.

5.6.2 Chromosomes and Genes Example

An example of the genetic algorithm's chromosomes is presented in Table 5.1 to explain chromosomes and genes of the genetic algorithm. Each chromosome will have n genes, each representing one of the n sensors to be placed in the distribution grid. The integer value stored in each of the n genes represents the node identifier number in the graph where that sensor is to be located. Table 5.1 shows an example population of four chromosomes and for optimizing the placement of 5 sensors. In this scenario E will hold the top 2 performing solutions. The fitness value column in Table 5.1 shows the amount of electrical cable of the grid that is categorized as PLE.

In this case chromosomes 2 and 4 would be placed in the elite set as placing sensors on the nodes in their chromosomes would result in the most PLE coverage (50Km and 42Km respectively). The selection process randomly chooses chromosomes from the elite set E and the population P (by using a random integer generator to generate the indices of the chromosomes) for the *crossoverAndMutate* function.

The *crossoverAndMutate* function creates a pair (parents) from the selected chromosomes in X. Each pair's genes (which are nodes selected as sensor locations) are used to generate a new chromosome. The number of chromosomes to be mutated is controlled by the value of *pm*. The value of *pm* is used as a percentage of the population size to determine the number of new chromosomes that will be mutated. If a new chromosome is to be mutated, then it is created by selecting random genes (by using a random integer generator) from all possible node locations. This is explained further in Section 5.6.3.

Table 5.1 An example population for the optimization problem

Solution	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Fitness Value (PLE Coverage)
1	215	79	139	107	7	16Km
2	33	220	214	90	8	50Km
3	10	174	40	109	9	6Km
4	218	176	103	64	28	42Km

After the mutations have been made a new generation of solutions would have been created. This population of new chromosomes would go through the same process of updating the elite set and creating new populations until $j = 50$ and the algorithm terminates. If a change in the elite set is detected j will be reset to 0 and the mutation rate would be reset to 0.02. When the while loop terminates the best solution in the elite set is then selected as the final solution.

5.6.3 Genetic Algorithm Support Functions

There are five functions used in the genetic algorithm:

1. Initialize
2. Evaluate
3. Update
4. Select
5. CrossoverAndMutate

The initialize function uses the number of sensors as the number of genes in the chromosomes and uses *pSize* as the number of chromosomes in the population. This function uses a random integer generator to create each chromosome's genes. The random generator is given the number of nodes available for sensor locations. Each node in the grid is identified by an integer from 1 (the first node) up to numberOfNodes (the last node).

Algorithm: Initialize

```
Input:  numberOfSensors, pSize
Output:  P //Initial population
P = new Array( pSize ,numberOfSensors)
for j=1: pSize do //generate each chromosome for the population
    for i=1: numberOfSensors  do //generate each chromosome's genes randomly
        P[i][j] = randi ( 1, numberOfNodes ) //random integer generator
    endfor
endfor
return P
```

The evaluation function used in the genetic algorithm is the Evaluate algorithm defined in Section 5.4. The update function is defined in the Update algorithm. The Update function maintains the elite set by updating it with any new high performing chromosomes (solutions). The update algorithm adds the elite set to the population and then sorts all chromosomes in descending order. The update function then takes the top ten solutions and overwrites the elite set with the best ten solutions.

Algorithm: Update

```
Input: performance, E, P
Output: E //Updated elite set
allSolutions = append ( P, E) //combine both elite solutions and population
Sort(allSolutions, performance) //sort using performance of each solution
for i=1: 10 do //only select top ten as elite set size is 10
        E[i][:] = allSolutions [i][:]
endfor
return E
```

The select function is shown in the Select algorithm. The select function takes the elite set E and the population P. A random integer generator is used to generate indices for choosing the solutions from the elite set E and then the population P that will be used in the crossover operation. The selected chromosomes from both E and P are placed in *X*.

Algorithm: Select

```
Input: E, P, numberOfSensors
Output: X //Set of chromosomes for crossover
for j=1: numberOfSensors do //select solutions from elite set randomly
    index = randi( 1, size( E ) )
    elite= E[index][:]
endfor
for i=1: numberOfSensors do //select solutions from population randomly
    index = randi( 1, size( P ) ) //random integer generator
    population = P[index][:]
endfor
X = append(elite, population ) //combine selected solutions for crossover
return X
```

The CrossoverAndMutate function takes the set X selected for crossover and mutation. It first calculates the value of *pm* (line 4) according to the rate of convergence. The mutateCount counter (line 5) ensures that the correct number of chromosomes are mutated according to the value of *pm*. The CrossoverAndMutate function then enters a loop (line 6 to 17) that creates enough chromosomes for a new population Two chromosomes are chosen from set X (line 7 and 8) as two parents are needed for each new member. The genes of both parents are then stored in genePool (line 9). Another for loop is then executed to create each gene for the new chromosome (line 10 to 16). If the chromosome is to be mutated (we have not mutated enough chromosomes yet), the genes are chosen randomly from all possible values which is any node in the grid (line 12). The *randi* function generates a random integer between the two parameters passed to the function. The *uniqueRandi* function takes the current chromosomes genes and the range of integers to generate. It generates a random integer that is not already in the chromosome, ensuring that the genes added to the chromosome are not repeated. If all of the mutated chromosomes have been created (line 11), then the genes are selected from the parent's genes in *genePool* (line 14). After creating all the new chromosomes, the new population is returned (line 18).

Algorithm: CrossoverAndMutate

```
1.  Input:  X, numberOfSensors, numberOfNodes , convergeCounter, pSize
2.  Output:  P //new population
3.  P = null
4.  pm = (convergeCounter * 0.02) //dynamically set pm based on convergence
5.  mutateCount = pm * pSize //used mutate a percentage of the new chromosomes
6.  for i=1: pSize do //create new chromosomes for new population
7.     parentA= X[ randi( 1, numberOfSensors) ][:] //select 1st parent at random
8.     parentB = X[ randi( 1, numberOfSensors) ][:] //select 2nd parent at random
9.     genePool = merge( parentA , parentB ) //collect genes and remove duplicates
10.    for j=1:  numberOfSensorsdo //create each gene for each new chromosome
11.      if (i < mutateCount) //decide if the chromosome's genes should be mutated
12.        P[i][j] = uniqueRandi ( P[i][:],  1, numberOfNodes )//random genes no duplicates
13.      else
14.        P[i][j] = genePool[ Randi( 1, numberOfSensors ) ] //use parent genes
15.      endif
16.    endfor
17. endfor
18. return P //return new generation
```

5.7 Optimization Algorithm Performance and Comparison

An initial view on the optimization methods' performance in relation to globally optimal solutions is now presented. This provides initial proof that the algorithms are capable of generating globally optimal solutions. As the optimization problem does not scale well, determining the globally optimal solutions using an exhaustive search was done on only placing two, three and four sensors on the 229-node test grid. The performance of the greedy algorithm in relation to the globally optimal solutions generated by a brute force approach is shown in table 5.1 where sensor node locations are the node IDs in the graph selected as sensor locations. The performance of the genetic algorithm in relation to the same globally optimal solutions is shown in table 5.2.

Table 5.2 Initial greedy algorithm results

Number of Sensors	Sensor node locations (Best)	Meters Covered by (Best)	Sensor node location (greedy)	Meters Covered by (greedy)
2	61 , 27	3816.7	61, 27	3816.7
3	61, 55, 12	6887.8	61, 55, 27	6871
4	61, 27, 12, 55	9008	61, 55, 27, 12	8908.7

The placement of the first two sensors was a globally optimal solution as expected as the *initialSensorPair* function is an exhaustive search. However, the greedy algorithm did not find the best solution for placing three and four sensors as seen in table 5.1. It must be noted however, that the difference in the greedy solutions and the globally optimal solutions are less than 1.5%. The greedy algorithm only provides near optimal solutions in these cases. These initial results show that the greedy algorithm may be an effective method for optimizing sensor locations but it may not be able to find globally optimal solutions.

Table 5.3 Initial genetic algorithm results

Number of Sensors	Sensor node locations (Best)	Meters Covered by (Best)	Sensor node locations (Genetic Algorithm)	Meters Covered by (Genetic Algorithm)
2	61, 27	3816.7	61, 27	3816.7
3	61, 55, 12	6887.8	61, 55, 12	6887.8
4	61, 27, 12, 55	9008	61, 55, 27, 12	9008

The results in table 5.2 show that the genetic algorithm is able to find the best possible solutions for all three test cases while the greedy algorithm does not. These results show that the genetic algorithm can provide the necessary flexibility needed to find solutions that the greedy algorithm cannot.

5.7.1 Performance Evaluation

The two algorithms provide different approaches to solving our optimization problem. The greedy algorithm sacrifices the exploration of the search space to provide solutions at a low computational cost. The genetic algorithm performs a deeper exploration of the large search space associated with this problem at the cost of additional computational resources. At this point we have established that the genetic algorithm is able to generate the optimal solutions for all three initial tests where the greedy algorithm was unable to do so.

We now perform a deeper evaluation on the two algorithms and provide a comparison between both methods. This main evaluation of the algorithms was done on placing varying amounts of sensors on a test grid provided by [Industry-Partner 2018]. Tests were carried out on this grid by each algorithm for the placement of two sensors up to ten sensors. These tests are performed on placing of two sensors up to ten sensors to first establish which method may be more effective. Additional tests are then performed in section 5.9 to expand the test cases for the selected solution to incorporate practical scenarios of the optimization method. The grid provided by [Industry-Partner 2018] contained 229 nodes. In the grid provided, each node is a random point in the electrical grid that allows the shape of the grid to be represented [Industry-Partner 2018]. Each node can be seen as a waypoint for the cables that make up the grid. The cost of placing the first two sensors is included in the results even though it will always be the globally optimal solution so that a complete comparison can be made between the algorithms.

The solutions generated for these cases by both methods were evaluated against each other. The testing is to establish the effectiveness of the greedy algorithm with a lower computational cost and a restrictive search and compare it with genetic algorithm that provides more flexibility at a higher computational cost. The results of this comparison will determine if the GA is able to generate better solutions by exploring more of the search space or if the greedy algorithm's is able to generate better solutions.

5.7.2 Resource Usage Performance

We now look at the resources used by each algorithm to gauge the efficiency of the two methods. We start with the greedy algorithm's time complexity. When placing the first two sensors an exhaustive search of all possibilities is done first. After the initial two sensors are placed the cost of each additional sensor is equivalent to a search of each unused node to determine the next best node. The time complexity is the sum of the cost placing the first two sensors and adding each additional sensor needed giving us $O(n!)$ where n is the number of nodes that can be used as locations for sensors in the grid.

To determine the number of iterations and resources needed by the GA we allow the genetic algorithm to run until the genetic algorithm has fully converged on a solution. However, a time complexity for the genetic algorithm cannot be expressed in terms of the number of nodes as its run time and resources vary with each run and are tied to the convergence of the algorithm. To compare the GA to the greedy algorithm we use the actual resources used by both methods expressed as the number of calls to the evaluation function. The data shown in figures 5.2, 5.3, 5.5 for the genetic and hybrid algorithm tests are averages taken from fifty tests for each number of sensors placed to account for the variability of the solutions that may be returned by these methods. The greedy algorithm data does not require averages, as it returns the same solutions for each number of sensors placed.

Figure 5.2 shows the difference in computational cost (expressed as the number of calls to the evaluate function) between the greedy algorithm and the genetic algorithm for optimizing sensor locations in the 229-node test grid.

The genetic algorithm uses much more resources on average in all cases (as expected) but the initial placement of two sensors. The genetic algorithm only requires 18% of the resources used by the exhaustive search for placing the initial two sensors. The quality of the solutions generated by each method is analyzed to determine how much of an increase in solution performance is attained from the additional computational cost of the genetic algorithm.

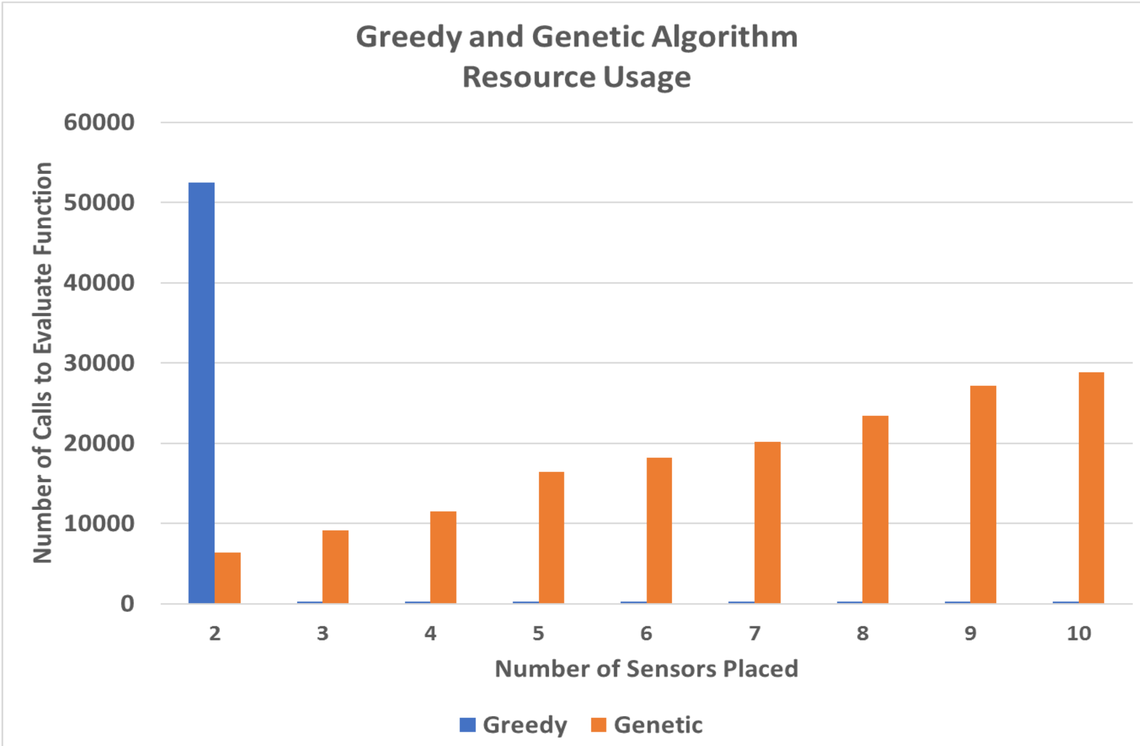

Figure 5.2 Resource usage for the greedy and genetic algorithms

5.7.3 Solution Performance

Figure 5.3 shows the amount of cable covered under primary localization by the greedy algorithm's solutions and the genetic algorithm's solutions for all of the tests performed.

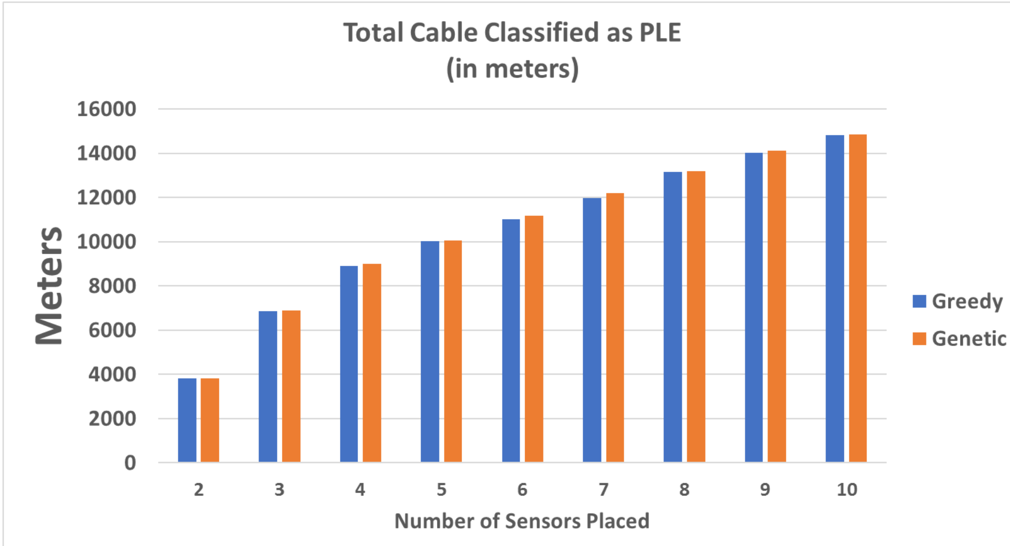

Figure 5.3 Meters of cable categorized as PLE

Figure 5.3 shows that the coverage gained by both methods is almost identical. This is a strong indicator that the greedy algorithm is more efficient at optimizing the location of sensors. In the initial tests the greedy algorithm did not find the absolute best solution for placing 3 or 4 sensors. However, the greedy algorithm only fell short by 16.8 and 99.3 meters (less than 1%) which may not have much impact when a line worker (someone who is trying to repair the fault) is looking for the fault location. Looking at all of the tests from placing two to ten sensors there is little to separate the two methods in final solution performance as seen in figure 5.3. The genetic algorithm was only able to increase performance by an average of around 0.6% an amount not visible in the data displayed in figure 5.3. However, the genetic algorithm's solutions were consistently better than the greedy algorithm's solutions. The genetic algorithm generated better solutions (though marginally) than the greedy algorithm.

5.8 Optimization Strategy

The results of our comparison are now used to establish a final optimization strategy for placing sensors in the grid. The greedy algorithm is more efficient in computational cost than the genetic algorithm for the tests performed. However, the greedy algorithm is not able to generate optimal solutions in some cases as seen in table 5.1. The greedy algorithm is limited by which solutions it can consider as each iteration chooses a sensor location that cannot be changed. It is possible that better solutions can be built by allowing a chosen sensor location to be changed.

5.8.1 The Hybrid Strategy

We propose the use of the greedy algorithm as the core optimization strategy as it has been proven to be very good at generating near optimal solutions with relatively little resources. We also propose the use of the genetic algorithm as a secondary optimization process that takes as input the solutions generated by the greedy algorithm. The genetic algorithm would take the less expensive greedy solution and place it in the initial population set of solutions used by the genetic algorithm effectively seeding the genetic algorithm with high performing genes. The genetic algorithm would apply the genetic operators to the population and run as normal. This would now allow the sensor locations in a greedy solution to be modified by mutation and crossover. From a genetic algorithm perspective, this would effectively widen the scope of the solution generation process to the global search space. This facilitates the modification of the greedy solution to potentially increase solution quality.

Algorithm 5: Hybrid

```
1.  Input:  numberOfSensors, numberOfNodes, Nodes
2.  Output:  locations
3.  greedySolution = Greedy(numberOfSensors)
4.  E = null //2D array to hold elite chromosomes
5.  convergeCounter = 0, max = 50, pSize = 100
6.  P = initialize( numberOfSensors , pSize, greedySolution ) //create initial population
7.  while  convergeCounter < max do //execute until the algorithm has converged
8.      for i = 1:size(P)  do //evaluate each chromosome (solution)
9.          performance(i) = Evaluate( P(i) ) //function as defined in section 5.4
10.     endfor
11.     prevE = E //temporarily store old elite set to check for change
12.     E = update(performance, E, P) //update E with best solutions in population
13.     if ( prevE = E ) //check for a change in elite set (detect convergence)
14.       convergeCounter = convergeCounter + 1
15.     else
16.       convergeCounter = 0 //reset counter whenever a change is detected
17.     endif
18.     X = select(E, P, numberOfSensors) //choose chromosomes for crossover
19.     P=crossoverAndMutate(X, convergeCounter, numberOfSensors, numberOfNodes)
20. endwhile
21. return locations = max( E ) //return best solution
```

There are two changes made to the original genetic algorithm (algorithm 4). These two changes are that the greedy algorithm (algorithm 3) is called in line 3 and the greedy solution generated is then passed to the initialize function (line 6). At line 6 the initialize function generates the initial population but inserts the greedy solution into the set. This allows the greedy solution to be modified so that all sensor locations chosen by the greedy method can be changed. Figure 5.4 shows the increase in the meters covered under PLE for the optimization solutions generated by the hybrid approach when compared to the greedy approach.

The hybrid algorithm is able to find better solutions than the greedy algorithm for every case except placing 2 sensors as this is the globally optimal solution. The hybrid algorithm has successfully modified the non-optimal greedy solutions to create the globally optimal solutions for placing 3 and 4 sensors (by comparing the hybrid solutions with the optimal solutions generated in the initial tests seen in table 5.1 and 5.2). However, it must be noted that the improvements at best were 180 meters. This equates to a 1.5% increase in performance for the biggest improvement. The hybrid approach allows better solutions to be generated but only marginally so.

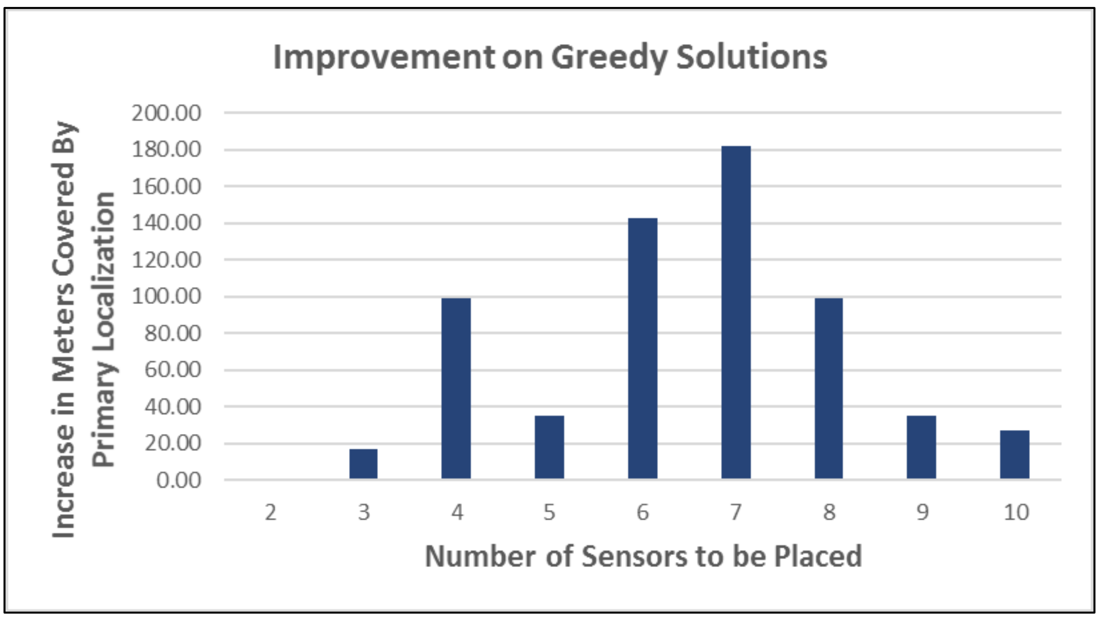

Figure 5.4 Improvements in solution performance over the greedy algorithm

Figure 5.5 shows the resource usage using the hybrid approach compared to the genetic and greedy algorithm. The hybrid algorithm's resource usage also includes the initial call to the greedy algorithm to generate the initial solution placed in the genetic algorithm's initial population. For placing the initial 2 sensors the genetic algorithm is the most effective, generating the globally optimal solution using less than 20% of the resources of the other methods. When placing the first 2 sensors the genetic algorithm could therefore be used. For the remaining problem sizes, the hybrid method can generate better solutions than the greedy method while using less resources than the genetic method. The hybrid method is successful in providing a more reliable method than the greedy approach as it is not limited by the iterative process while being more efficient than the genetic algorithm.

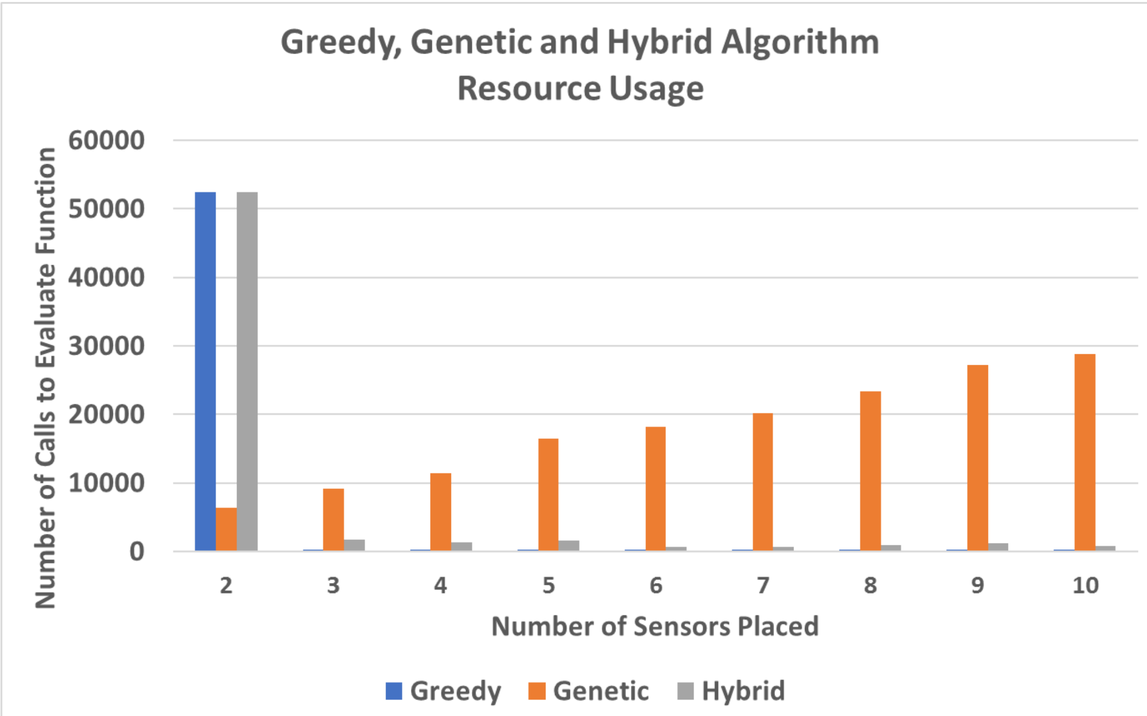

Figure 5.5 Resource usage for Greedy, Genetic and Hybrid methods

5.9 Considerations for Practical Applications

The hybrid method provides a reliable and efficient way of calculating the best locations to place n sensors in an electrical grid to maximize PLE coverage. The primary consideration for the application of such a method is choosing the number sensors (choosing n) that should be placed in the electrical grid for any given scenario. Additional tests were performed using the hybrid method for placing up to fifty sensors on the test grid. The tests were done to provide a more complete view of the coverage that could be attained for a larger number of sensors which may provide insight into selecting a value for n (the number of sensors) in a given scenario. The PLE coverage for these tests is shown in figure 5.6.

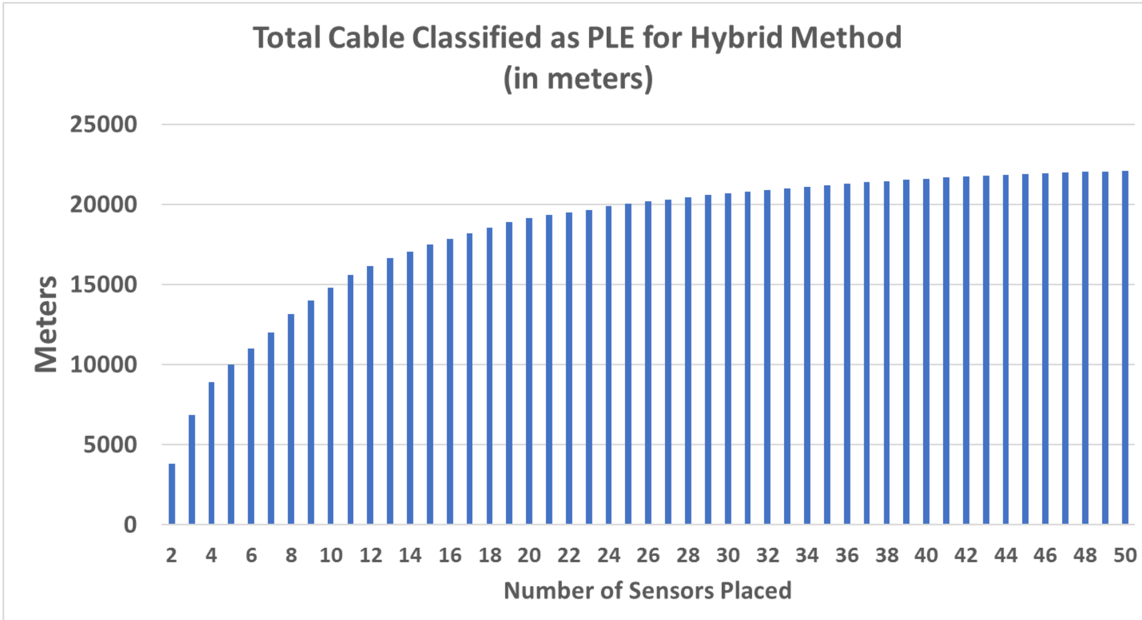

Figure 5.6 Total cable classified as PLE for the Hybrid algorithm

Figure 5.7 shows the amount of PLE coverage gained for each additional sensor. This coverage per sensor metric provides a useful perspective of the optimization results. Figure 5.7 shows that there is a significant drop in additional coverage as more sensors are added. The total cable in the test grid is 23 kilometers. The amount of coverage for adding the first 5 sensors is 43% (over 10 kilometers). However, the next 5 sensors (10 total) only provide 20% more coverage. After placing another 5 sensors (15 total) the increase in PLE coverage has fallen to 1%. The coverage added per sensor consistently declines as more sensors are added. The question formed from these results is at what point is it no longer beneficial from a cost-based perspective. This is a particularly interesting perspective and is a strong indicator that there is no single value for n (the number of sensors to place in the grid) that would work in all cases. This is because each implementation of the fault localization solution may require n (the number of sensors to place in the grid) to be calculated based on the needs and standards of the company using the solution.

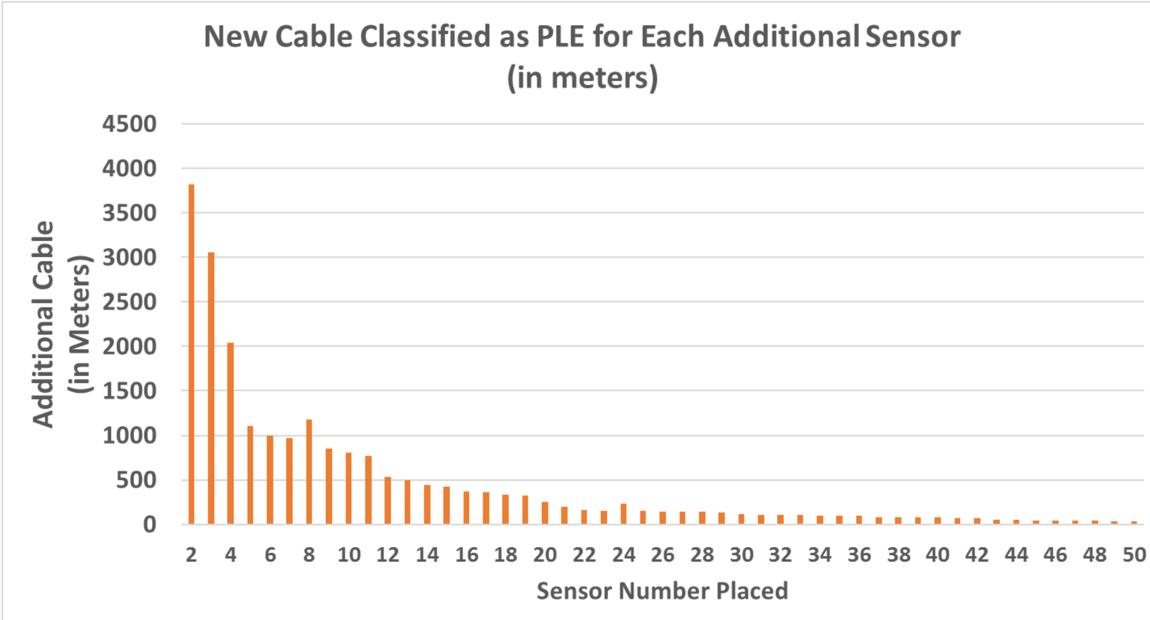

Figure 5.7 Amount of new cable classified as PLE for adding each sensor

The test data analysis indicates that the selection of n is not a simple process and requires specific information and decision making from the user employing the fault localization solution. The hybrid optimization algorithm does however provide a means for assisting the process of selecting the number of sensors to place in the grid.

- The hybrid optimization algorithm can be used to generate a coverage per sensor metric that can be used with a threshold (of meters) so that once adding an n^{th} sensor does not add enough PLE coverage then no more sensors are placed.
- Similarly to the first point, the coverage per sensor metric can also be used to establish a cost per meter of coverage for each sensor using the cost of a sensor and the PLE coverage gained for each sensor. This can also be used in conjunction with some threshold to ensure the cost per meter coverage does not exceed a certain amount.
- The optimization algorithm can also be used to provide a business with an estimate of the coverage for a particular number of sensors that may be set based on a budget. If a budget is provided, the total PLE coverage for that budget can be calculated by running the optimization algorithm using the number of sensors the budget can support as the value for n.

- It is also possible that the number of sensors can be determined by the desired coverage under the PLE. Let us say for example that the electric company wants 80% of the grid to be covered under the primary localization scheme. In this scenario the value of n is increased until the desired percentage is covered under the primary localization scheme. By doing this an adequate number of sensors (value for n) and the locations where this coverage can be achieved can be calculated.

These examples for the application of the optimization algorithm is not an exhaustive list. More ways of determining n can be derived from any number of characteristics present in the business model of a company employing the optimization algorithm.

5.10 Conclusion and Future Work

This chapter builds on a previously discussed method of localizing faults. This method categorized areas of the grid as being covered under the primary or secondary localization scheme. This chapter looks at the development of a method to place sensors in an electrical distribution grid to cover as many meters of cable as possible under the primary localization scheme as it is the most accurate fault localization scheme. Initially, two methods were developed and then evaluated. The first was a simple greedy method that incrementally chooses the best new sensor location until the desired number of sensors were placed. The second was a genetic algorithm that would perform a more robust evaluation of possible solutions. The performance of these two methods showed the greedy approach in most cases was as good as the genetic algorithm while using significantly less resources. This would have been the chosen method but for the fact that the greedy algorithm cannot handle all cases due to the incremental nature of its approach. The greedy algorithm was unable to handle cases where a better solution may be found by changing the sensor locations of previous iterations. To address this issue a hybrid approach was developed that used the greedy approach to generate the initial solution and then applied the genetic algorithm to allow the solution to be further modified. The hybrid approach was then able to provide a much more efficient optimization strategy. We acknowledge that in environments where resources are very limited, the greedy algorithm can act as an effective optimization strategy for the optimization problem. For cases where there is not such a strict resource constraint the hybrid approach can be used to ensure solution quality. We now highlight considerations and notable areas for future work associated with the optimization algorithm development performed in this chapter.

5.10.1 Limits of Graphical Representations of Electrical Grids

A limitation of this work is created because the optimization algorithm takes a set of nodes and edges that represent the electrical grid and selects sensor locations from that set of nodes. Traditionally, the nodes of the graphical representation of a grid are the busses of that electrical grid. A bus refers to a location where the grid branches off into different directions. The optimization algorithm in this case would be choosing a set of busses on which the sensors would be placed. This was seen a restricted approach to determining sensor locations as sensors in practice can be placed anywhere on the power cables. For this work we wished to acknowledge that sensors could be placed in locations other than busses. For the work presented in this chapter, the set of nodes for the grid provided by our industry partner for our tests represented the busses and other (random) locations along the cables between those busses. This grid was therefore able to allow our optimization algorithm to consider locations between busses as possible locations which aligns with our goal of acknowledging sensor locations may be non-bus locations. The work presented in this chapter shows that the algorithm can optimize the locations of sensors not only at busses but also when the possible locations are anywhere in the electrical grid. There is still work to be done in this area as there is a need to perhaps create an optimization algorithm that does not need nodes to be provided. A next step for this work therefore, could be to find a way to incorporate the edges of a graph into the optimization process.

5.10.2 Prioritized Monitoring

Another avenue for further development is in the use of historical data for prioritized monitoring. Access to historical data on the age of components (like transformers etc.) and the locations of previous faults, would allow high risk areas of the grid to be identified. These high-risk areas could potentially be prioritized for monitoring by adjusting the weight of the edge in the graph that corresponds to that high-risk area. The use of some metric would need to be established, which would be added to the weight of the edge. This in theory, could allow the optimization algorithm to prioritize these edges resulting in an increased effort to cover the high-risk areas under the primary localization scheme.

5.11 Summary

The optimization algorithms presented in this paper are the final contribution to research in the area of fault localization in electrical distribution grids. These algorithms are seen as an initial look at the effect that the locations and number of sensors present on the grid has on the ability to localize faults. In this chapter, a simple greedy method's ability to place sensors in the grid is compared to a more resource intensive genetic algorithm. The goal of these algorithms is to maximize the area covered under the highly accurate primary localization scheme. The work in this chapter lays the groundwork for evaluating different optimization approaches for the fault localization solution presented in Chapter 4. The choice of optimization criteria must still be developed based on industry needs, which prevents a final decision on what optimization approach would be best to employ. We therefore created these algorithms so that they can act as a framework that be easily modified to incorporate different constraints and even different objectives if needed. The greedy, genetic and hybrid algorithms presented use a single evaluation function. The evaluation function can easily be replaced by one that optimizes goals provided by industry. Some examples of these alternate goals are discussed at the end of Section 5.9. Chapter 6 draws conclusions on the three main components of work presented in this dissertation (synchronization, fault localization and optimization).

5.12 Bibliography

Bandyopadhyay, S. and S. Saha. Some Single- and Multiobjective Optimization Techniques. Unsupervised Classification, Springer Berlin Heidelberg**:** 17-58, (2013).

Barron, A. R., A. Cohen, W. Dahmen and R. A. DeVore. "Approximation and learning by greedy algorithms." *The annals of statistics* **36**(1): 64-94, (2008).

Cormen, T. H., C. E. Leiserson, R. L. Rivest and C. Stein. Introduction to algorithms, MIT press, (2009).

Crossley, P. A. and P. G. McLaren. "Distance Protection Based on Travelling Waves." *IEEE Transactions on Power Apparatus and Systems* **PAS-102**(9): 2971-2983, (1983).

Das, S., S. Kulkarni, N. Karnik and S. Santoso. Distribution fault location using short-circuit fault current profile approach. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Dereventsov, A. V. "Comparison of the convergence rate of pure greedy and orthogonal greedy algorithms." *Mathematical Notes* **92**(3): 485-489, (2012).

DeVore, R. A. and V. N. Temlyakov. "Some remarks on greedy algorithms." *Advances in Computational Mathematics* **5**(1): 173-187, (1996).

Hamo, Y. and S. Markovitch. The COMPSET Algorithm for Subset Selection. *IJCAI*, (2005).

Hizam, H. and P. A. Crossley. Single Ended Fault Location Technique On a Radial Distribution Network Using Fault Generated Current Signals, (2006).

Hunte, J., H. Lutfiyya and A. Haque. Fault Localization in Smart Grids Using Segmentation. *ICC 2021 - IEEE International Conference on Communications*, (2021).

Industry-Partner. Ontario Electricity Provider, (2018).

Karnik, N., S. Das, S. Kulkarni and S. Santoso. Effect of load current on fault location estimates of impedance-based methods. *2011 IEEE Power and Energy Society General Meeting*, (2011).

Lin, S., Y. Rong, X. Sun and Z. Xu. "Learning Capability of Relaxed Greedy Algorithms." *IEEE Transactions on Neural Networks and Learning Systems* **24**(10): 1598-1608, (2013).

Magnago, F. H. and A. Abur. "Fault location using wavelets." *IEEE Transactions on Power Delivery* **13**(4): 1475-1480, (1998).

Qian, C., J.-C. Shi, Y. Yu and K. Tang. On Subset Selection with General Cost Constraints. *IJCAI*, (2017).

Saha, M., F. Provoost and E. Rosolowski. "Fault location method for MV cable network." (2001).

Sapountzoglou, N., B. Raison and N. Silva. Fault Detection and Localization in LV Smart Grids. *2019 IEEE Milan PowerTech*, (2019).

Škumát, T. and E. Ž. Open Phase Fault Localization in a Medium Voltage Distribution Grid Using Smart Meters. *2019 20th International Scientific Conference on Electric Power Engineering (EPE)*, (2019).

Srinivas, M. and L. M. Patnaik. "Adaptive probabilities of crossover and mutation in genetic algorithms." *IEEE Transactions on Systems, Man, and Cybernetics* **24**(4): 656-667, (1994).

Srinivas, M. and L. M. Patnaik. "Genetic algorithms: a survey." *Computer* **27**(6): 17-26, (1994).

Takagi, T., Y. Yamakoshi, M. Yamaura, R. Kondow and T. Matsushima. "Development of a New Type Fault Locator Using the One-Terminal Voltage and Current Data." *IEEE Transactions on Power Apparatus and Systems* **PAS-101**(8): 2892-2898, (1982).

Temlyakov. "Nonlinear Methods of Approximation." *Foundations of Computational Mathematics* **3**(1): 33-107, (2003).

Trindade, F. C. L. and W. Freitas. "Low Voltage Zones to Support Fault Location in Distribution Systems With Smart Meters." *IEEE Transactions on Smart Grid* **8**(6): 2765-2774, (2017).

Trindade, F. C. L., W. Freitas and J. C. M. Vieira. "Fault Location in Distribution Systems Based on Smart Feeder Meters." *IEEE Transactions on Power Delivery* **29**(1): 251-260, (2014).

Vafaie, H. and I. F. Imam. Feature selection methods: genetic algorithms vs. greedy-like search. *Proceedings of the international conference on fuzzy and intelligent control systems*, (1994).

Chapter 6

6 Conclusions and Future Work

The main goal of the research presented in this dissertation was to develop a fault localization method that aligns with the needs of power distribution companies and provide movement towards Smart Grid development as outlined in the NIST Smart Grid Interoperability Standards Roadmap [Gopstein et al. 2021]. The distribution grid is a difficult environment for fault localization [Jun et al. 1997]. The method developed is an extension of the existing work in fault localization that address two main limitations found in existing work. These limitations were found to be:

- the use of complex data that may not be available or may only be available in some scenarios
- the general assumption of synchronized data across all devices used to generate data

To address these limitations three goals were established under the following research headings:

- Synchronization of Multiple Devices – This refers to an alternate technique for synchronizing the data captured across multiple devices that may be present in the electrical grid
- Localization Using Segmentation and Multiple Devices – This refers to the development of a new solution for localizing faults on the distribution grid
- Optimization of Device Locations – This refers to the creation of an optimization method for determining the most effective location for the sensors used in the new fault localization technique

In an effort to satisfy these goals, three major contributions were made to fill open areas of research. First a novel synchronization method was developed to remove the assumption of synchronized data and reliance on a pre-existing synchronization scheme. Second, a new localization method was developed that provides highly accurate fault localization in the distribution grid using multiple devices. Thirdly an optimization algorithm was created that determines the most effective locations for sensors for the new fault localization scheme.

6.1 Synopsis of Results

We present a synchronization method in Chapter 3 that shows much promise in allowing faults to be localized by using more than one device. This work facilitates the development of a more robust localization system as reliance is no longer placed on the accuracy of a single measurement from a single device. Furthermore, by removing the need for expensive clocks or components to maintain synchronization of the devices and their data is not needed, which adds to the robustness and simplicity of the solution while reducing the cost of implementation.

The work built on this initial synchronization step in Chapter 4 provides an approach to fault localization in distribution grids that uses multiple sensors, segmentation and transient signals caused by a fault. This method can determine the precise location of the fault in scenarios where many possible locations for the fault exist. Our evaluations show that the approach is successful based on results that show that the fault localization solution can accurately localize faults on a distribution grid.

To offer a complete solution, an optimization method was developed. This method allows the optimization of devices used by the fault localization method. The algorithm takes as input a graphical representation of the grid and the desired number of devices. The optimization algorithm then determines the best locations for those sensors. The goal was to maximize the amount of cable or number of edges covered under the primary localization scheme introduced in [Hunte et al. 2021]. The greedy algorithm generated solutions almost as good as the more robust genetic algorithm. The genetic algorithm consistently created better solutions. As a result of the similarities on solution performance, a greedy algorithm has been identified as a possible option for the optimization problem when resources are limited. However, a hybrid approach using greedy and genetic algorithm components was chosen as the best option given it has been shown to outperform the greedy algorithm while using significantly less resources than the genetic algorithm.

6.2 Future Work

One of the most pressing avenues for future work is in the development of test grids that are designed for transient analysis. This remains a pressing issue as many test grids have been developed for other purposes, like power flow analysis. As such the transient signals and data present during faults and the key to many fault localization solutions is not a primary objective of these test grids. In this regard, an important factor to note is that the speed of these transient signals requires a very high sample rate. The high sample rate would result in large quantities of data being generated for each simulated component on the test grids. The ability to create realistic test grids with efficient methods of capturing data needed for transient analysis therefore poses quite a challenging problem. The existence of a standardized set of test grids would provide a consistent foundation for fault localization research to be conducted. Some of the supplemental avenues for future work linked specifically to the work presented in this dissertation are listed as follows:

- The augmentation of the fault localization solution to possibly increase the accuracy of the solution

- Establishing a globally accepted metric for optimization of sensors in the industry for fault localization.

- Implementation of the entire work presented in this dissertation in an active electrical grid.

Sections 6.2.1 to 6.2.3 describe these avenues for future work in more detail.

6.2.1 Historical and Geographical Supplemental Data

Another area for development is in the use of historical and geographical data in the localization process to augment the fault localization solution presented in this dissertation. This would require a detailed relationship between faults and several characteristics of the grid's components and their immediate environment to be established. This also leads into a possible area for future development with the optimization of sensor locations. The identification of high-risk areas of the grid can then allow such areas to become a priority for monitoring by the optimization algorithm.

6.2.2 Alternate Optimization Metrics

The optimization work presented in this dissertation successfully identifies an effective method of determining effective location for sensors in the distribution grid for fault localization. There is more to explore in this area, specifically with the metrics that are used to assess possible solutions to the optimization problem. As discussed in section 5.9, there are a number of options that may influence what is considered an acceptable solution. There is therefore a need to initiate a conversation with industry experts to determine which factors are considered the most effective when it comes to the sensors used by the fault localization solution.

6.2.3 Additional Testing

The work presented has attracted significant attention in the power industry. A number of power distribution companies have indicated a keen interest in the solution developed for implementation in their distribution grids. In response, a provisional patent has been filed for the presented work. The next steps for the work in relation to this are focused on providing the results of final testing required to convert the provisional patent into a full patent. These tests are to implement the synchronization, localization and optimization components presented in a real-world test grid rather than simulations. This integration would first require a discussion with the industry to identify existing systems that may be leveraged to integrate the fault localization solution with the existing business practices. These integrations and tests would ultimately convert the success of the simulated tests into a proven and implemented solution.

6.3 Bibliography

Gopstein, A., C. Nguyen, C. O'Fallon, N. Hastings and D. Wollman. "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 4.0." *NIST Spec. Publ*, (2021).

Hunte, J., H. Lutfiyya and A. Haque. Fault Localization in Smart Grids Using Segmentation. *ICC 2021 - IEEE International Conference on Communications*, (2021).

Appendix Copyright Permissions

Appendix A: International Conference on Communications 2021 Copyright

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

Fault Localization in Smart Grids Using Segmentation
JACOB HUNTE,Hanan Lutfiyya,Anwar Haque
ICC 2021 - IEEE International Conference on Communications

COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event ofa breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the IEEE PSPB Operations Manual.
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance bythe IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

You have indicated that you DO NOT wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

[redacted]

Signature

07-03-2021

Date (dd-mm-yyyy)

Information for Authors

AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEEs publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

AUTHOR ONLINE USE

- **Personal Servers**. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
- **Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, ereserves, conference presentations, or in-house training courses.
- **Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

Questions about the submission of the form or manuscript must be sent to the publication's editor.
Please direct all questions about IEEE copyright policy to:
IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966

Appendix B: IEEE 7th World Forum on Internet of Things 2021 Copyright

IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

Device Synchronization for Fault Localization in Electrical Distribution Grids
Jacob Hunte, Anwar Haque, Hanan Lutfiyya
2021 IEEE 7th World Forum on Internet of Things (WF-IoT)

COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the IEEE PSPB Operations Manual.
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

You have indicated that you DO wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."

CONSENT AND RELEASE

1. In the event the author makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the author, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.
2. In connection with the permission granted in Section 1, the author hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement,

1. distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

[redacted]

Signature

07-05-2021

Date (dd-mm-yyyy)

Information for Authors

AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

AUTHOR ONLINE USE

- **Personal Servers**. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
- **Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, ereserves, conference presentations, or in-house training courses.

- **Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

Questions about the submission of the form or manuscript must be sent to the publication's editor.
Please direct all questions about IEEE copyright policy to:
IEEE Intellectual Property Rights Office, copyrights@ieee.org, +1-732-562-3966

Curriculum Vitae

Name: Jacob Hunte

Education:

University of Western Ontario
London, Ontario, Canada
PhD. Computer Science (Candidate), December 2021

The University of the West Indies
Barbados, West Indies
M.Sc. Computer Science, December 2015

The University of The West Indies
Barbados, West Indies September
B.Sc (Hons) Computer Science (Double), July 2012

Barbados Community College
Barbados, West Indies
A.A.B Business Studies May 2008

Honors and Awards:

Western Graduate Research Scholarship (WGRS) 2016 – 2020
Emerging Leaders in the Americas Program scholarship (ELAP) 2015
Sol Petroleum Caribbean scholarship for master's research. 2014 - 2015
Systems Consulting Limited Prize in Computer Science. 2011 – 2012
Kurleigh King Memorial Scholarship. 2011- 2012

Relevant Experience:

Course Instructor, Department of Computer Science
University of Western Ontario, London, Ontario, Canada
September 2020 – April 2021.

Graduate Research and Teaching Assistant, Department of Computer Science
University of Western Ontario, London, Ontario, Canada
September 2016 – Present

Developer, Business Systems
Cromwell Tools Ltd, Wigston, Leicester, England
February 2016 – August 2016

Web and Artificial Intelligence System Development Research and Teaching Assistant
The University of the West Indies, Barbados, West Indies
September 2013 –December 2015

Systems Developer
CIBC First Caribbean International Bank, Barbados, West Indies
June 2012 – April 2013

Systems Developer (Internship)
CIBC First Caribbean International Bank, Barbados, West Indies
May 2011 – September 2011

Systems Developer (Internship)
CIBC First Caribbean International Bank, Barbados, West Indies
June 2010 – September 2010

Business Assistant (Internship)
Colonial Life Insurance Company Ltd (CLICO), Barbados, West Indies
June 2007 - Aug 2007

Publications:

Improving genetic algorithm solution performance for optimal order allocation in an e-market with the Pareto-optimal set.
Nature-Inspired Computation and Swarm Intelligence. April 2020.
Authors: Hunte, J., Gittens, M., Gittens, C., Lutfiyya, H., & Edward, T

Fault Localization in Smart Grids Using Segmentation.
ICC 2021-IEEE International Conference on Communications June 2021.
Authors: Hunte, J., Lutfiyya, H., & Haque, A.

Device Synchronization for Fault Localization in Electrical Distribution Grids.
WF-IoT IEEE 7th World Forum on Internet of Things. June 2021.
Authors: Hunte, J., Lutfiyya, H., & Haque, A.