FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
ELECTRICAL ENGINEERING GRADUATE PROGRAM

WORK TITLE

# Clustering-Driven Equipment Deployment Planner and Analyzer for Wireless Non-Mobile Networks Applied to Smart Grid Scenarios

AUTHOR'S NAME

**Ladislav Vrbský**

DM 13/2018

**UFPA / ITEC / PPGEE**
**Guamá University Campus**
**Belém-Pará-Brazil**
**2018**

FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
ELECTRICAL ENGINEERING GRADUATE PROGRAM

AUTHOR'S NAME

**Ladislav Vrbský**

WORK TITLE

# Clustering-Driven Equipment Deployment Planner and Analyzer for Wireless Non-Mobile Networks Applied to Smart Grid Scenarios

DM 13/2018

**UFPA / ITEC / PPGEE**
**Guamá University Campus**
**Belém-Pará-Brazil**
**2018**

FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
ELECTRICAL ENGINEERING GRADUATE PROGRAM

AUTHOR'S NAME

**Ladislav Vrbský**

WORK TITLE

# Clustering-Driven Equipment Deployment Planner and Analyzer for Wireless Non-Mobile Networks Applied to Smart Grid Scenarios

Master's Thesis submitted for evaluation of the Examination Committee of Electrical Engineering Graduate Program of UFPA as a requirement for obtainment of title Master of Electrical Engineering in the area of Applied Computing.

**UFPA / ITEC / PPGEE**
**Guamá University Campus**
**Belém-Pará-Brazil**
**2018**

Dados Internacionais de Catalogação - na – Publicação (CIP)
Sistema de Bibliotecas da UFPA

---

Vrbský, Ladislav, 1992-
Clustering-driven equipment deployment planner and analyzer for wireless non-mobile networks applied to smart grid scenarios / Ladislav Vrbský. - 2018.
Orientador: Marcelino Silva da Silva
Dissertação (mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2018.

1. Cluster (sistema de computador) – análise. 2. Mineração de dados (computação). 3. Comunicação sem fio – modelos matemáticos. I. Título.

CDD 23. ed. 004.35

---

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

# "CLUSTERING-DRIVEN EQUIPMENT DEPLOYMENT PLANNER AND ANALYZER FOR WIRELESS NON-MOBILE NETWORKS APPLIED TO SMART GRID SCENARIOS"

AUTOR: **LADISLAV VRBSKY**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 16/03/2018

**BANCA EXAMINADORA:**

Marcelino Silva da Silva
**Prof. Dr. Marcelino Silva da Silva**
(Orientador – PPGEE/UFPA)

Diego Lisboa Cardoso
**Prof. Dr. Diego Lisboa Cardoso**
(Avaliador Interno – PPGEE/UFPA)

**Prof. Dr. Bruno Souza Lyra Castro**
(Avaliador Externo ao Programa – CCAST/UFPA)

**VISTO**:

**Prof.ª Dr.ª Maria Emília de Lima Tostes**
(Coordenadora do PPGEE/ITEC/UFPA)

*"Waste no more time arguing about what a good man should be. Be one."*

(Marcus Aurelius)

# Acknowledgments

First and foremost, to my amazing parents, who selflessly supported me with all they got, guided me through life as they know best, and gave me the opportunity to focus on the important, such as my education. To this I am very grateful and having such solid base is what enabled me to focus.

To my wife and our daughter, who both made me a better, more focused and driven. My wife's support was essential for a successful completion of this work. Having a family to take care of helped me to push through hard times and tiresome work, and to withstand more.

To all remaining family members, both in Czech and Brazil, for all kinds of support, local or remote, and the faith they have had in me over the many years.

To friends I made during my life wherever I went, who brightened my days with lots of fun and memories I will always remember.

To sports, especially baseball, where I found much joy. To coaches and players I met, and games I played, who made me a good team player, taught me lessons I will cherish for the rest of my life.

To professors and fellow researchers at labs LPO and LPRAD, who helped me to overcome the language barrier, as well as to understand new concepts and ideas, and created a lab environment pleasing to work in.

To educational systems, professors, and advisors of Brazil, USA, and especially Czech Republic, that made me an educated man generally, as well as in my fields of focus. They provided high-quality education and taught me persistence, resilience, and others.

To CNPq for the financial support that was provided to me, even though I am a foreigner. This aid gave me incentives to focus on my studies in Brazil, which led to a successful completion of the Master's education period.

Finally, to God, who has kept challenging me for many years, which made me develop a good amount of resilience, and who also makes life interesting and worth living.

# Contents

# List of Figures

# List of Tables

# Glossary

| | |
|---|---|
| 3GPP | Third Generation Partnership Project |
| AAM | Advanced Asset Management |
| ADO | Advanced Distribution Operation |
| AI | Artificial Intelligence |
| AHC | Agglomerative Hierarchical Clustering |
| AMI | Advanced Metering Infrastructure |
| ANSI | American National Standards Institute |
| AP | Access Point |
| ATO | Advanced Transmission Operation |
| BS | Base Station |
| CAPEX | Capital Expenditure |
| CH | Cluster Head |
| CIGRE | International Council on Large Electric Systems |
| D&D | Design and Development |
| DECC | UK Department of Energy and Climate Change |
| DNP3 | Distributed Network Protocol 3.0 |
| DOE | US Department of Energy |

| | |
|---|---|
| ETP | European Technology Platform |
| GAT | Generate and Test |
| GDP | Gross Domestic Product |
| GS | Greedy Search |
| HAN | Home Area Network |
| IDCA | Intelligent Deployment and clustering Algorithm |
| IEC | International Electro technical Commission |
| IED | Intelligent Electronic Device |
| IEEE | Institute of Electrical and Electronics Engineers |
| IP | Internet Protocol |
| IPv6 | Internet Protocol version 6 |
| ISO | International Organization for Standardization |
| ITU | International Telecommunication Union |
| JISC | Joint Information Systems Committee |
| KATS | Korean Agency for Technology and Standards |
| ML | Machine Learning |
| NIST | National Institute of Standards and Technology |
| OFDMA | Orthogonal Frequency-Division Multiple Access |
| OPEX | Operating Expenditure |
| PAM | Partitioning Around Medoids |
| PCA | Principal Component Analysis |
| PES | Power & Energy Society™ |

| | |
|---|---|
| PICOM | Piece of Information for Communication |
| PLC | Power Line Communication |
| QoS | Quality of Service |
| R&D | Research and Development |
| RTP | Real Time Protocol |
| SA | Simulated Annealing |
| SCADA | Supervisory Control and Data Acquisition |
| SF-SSCC | CEN-CENELEC-ETSI Sector Forum on Smart and Sustainable Cities and Communities |
| SINR | Signal-to-interference-plus-noise ratio |
| SVD | Singular Value Decomposition |
| SVM | Support Vector Machines |
| TDD | Time Division Duplex |
| UDP | User Datagram Protocol |
| UMTS | Universal Mobile Telecommunications System |
| UTRA-TDD | UMTS Terrestrial Radio Access - Time Divison Duplexing |
| WiMAX | Worldwide Interoperability for Microwave Access |
| WSN | Wireless Sensor Network |

# Abstract

The modern power grids, known as smart grids, rely on various advancements, one of them being the introduction of bi-directional communication. In some cases, data exchanged in the network is of critical importance. The data transmissions need to meet specific delay limits set by the regulatory agencies in order for the smart grid to function properly. Meeting these standards allows the use of new applications of monitoring, control and system protection, resulting in a more efficient, stable and environment-friendly system. This thesis presents a methodology for analysis and planning of wireless communication networks for smart grid, which uses a clustering algorithm to determine the optimal positions of the routers and gateways of the network to be installed. After, it calculates the delay for each Intelligent Eletronic Device that is a network subscriber. This way, an analysis can be made to obtain the Quality of Service requested for a specific network setup in a specific scenario. The results obtained in the performed case study show that it is possible to achieve a network topology that satisfies the maximum delay requirements of 100 % of its subscribers, using WiMAX or a combination of Wi-Fi and WiMAX. Also, the thesis explores a restricted communication mode that can temporarily suspend the transferences of non-critical data. In most scenario configurations, the restricted mode delivers all the data within the maximum delay. The software implementation of the proposed model is made publicly available under open-source license, so that anyone, including researchers, or private and public companies, can take advantage of it. The model presented in this thesis is customizable, allowing the use of other technologies and be used with other networks, including scenarios that are not related to smart grid.

**KEYWORDS:** Clustering Algorithms; Quality of Service; Smart Grid; Communication Networks; Communication Network Planning.

# Resumo

Os modernos sistemas elétricos de potência, conhecidos como *smart grids*, contam com vários avanços, sendo um deles a introdução de comunicação bidirecional. Em alguns casos, os dados trocados na rede são de importância crítica. As transmissões de dados devem atender aos limites de atraso específicos estabelecidos pelas agências reguladoras para que o *smart grid* funcione corretamente. O cumprimento desses padrões permite o uso de novas aplicações de monitoramento, controle e proteção do sistema, que resultam em um sistema elétrico mais eficiente, estável e ecológico. Esta tese apresenta uma metodologia para análise e planejamento de redes de comunicação sem fio para *smart grid* que usa um algoritmo de clusterização para determinar as posições ideais dos pontos de acesso e *gateways* da rede a serem instalados. Depois, calcula o atraso para cada dispositivo eletrônico inteligente que atua como o assinante da rede. Desta forma, uma análise pode ser feita para obter a Qualidade de Serviço almejada para uma determinada configuração de rede específica em um cenário específico. Os resultados obtidos no estudo de caso realizado mostram, que é possível alcançar uma topologia de rede, que satisfaça os requisitos de atraso máximo de 100 % dos seus assinantes, usando WiMAX ou uma combinação de Wi-Fi e WiMAX. Além disso, a tese explora um modo de comunicação restrito que pode suspender temporariamente as transferências de dados não críticos. Na maioria das configurações de cenário, o modo restrito entrega todos os dados dentro do prazo máximo. A implementação do software do modelo proposto é disponibilizada publicamente sob licença *open-source*, para que qualquer pessoa, incluindo pesquisadores, ou empresas privadas e públicas, possa aproveitá-lo. O modelo apresentado nesta dissertação é customizável, permitindo o uso de outras tecnologias e ser usado com outras redes, inclusive para cenários que não são relacionados ao *smart grid*.

**PALAVRAS-CHAVE:** Algoritmo de Clusterização; Qualidade de Serviço; *Smart Grid*; Redes de Comunicação; Planejamento de Redes de Comunicação.

# Chapter 1

# Introduction

## 1.1 Motivation and Problem Definition

Electricity has long been a major driving force for development of the modern world, boosting world economy and innovation. Until today, however, fossil fuels still remain the major energy source [DOE, 2018; WWF, 2010]. Rising concerns on greenhouse gas emissions, increasing electricity prices and diminishing fossil fuels have also called for the modernizations of the power grid [Erol-Kantarci and Mouftah, 2015]. According to data from [World Bank, 2017], the overall global dependency on fossil fuels in all of energy consumption (not restricted to electricity) has been stagnant at 81 % between 1990 and 2013. Further, global CO2 emissions rose 60 % over the same period. This is mainly caused by growing countries, where major focus in regards to energy is its cost. Further, the overall energy consumption has been steadily increasing, meaning the gross amount of fossil fuel consumption has been growing as well.

Companies and governments alike push the transition towards renewable energy usage. As an example of the clean energy initiative, Google Inc. has recently reported on twitter[1] about its investments in renewable energy sources, which bring the total capacity of the company to over 3 gigawatts, which is about as much as all of the company's services used in 2017.

[1]https://twitter.com/Google/status/936279094605504513

A modern approach to the existing power grid and the electricity lifecycle itself, called smart grid, has been one of the key research topics of in the energy sector, especially in the past decade since its first appearance in an official document, the Energy Independence and Security Act of 2007 [EISA, 2007], released by the United States Congress. The vision of smart grid is a general evolution of the whole electrical grid towards more stable, efficient, and environmentally friendly system.

The importance of smart grid is also visible from the financial perspective. National Institute of Standards and Technology (NIST) reports about the The US investments in [NIST 3.0, 2014], that every US$ 1 million of direct spending into the smart grid sector increased the Gross Domestic Product (GDP) by US$ 2.5 to US$ 2.6 million. Since the GDP multiplier is higher than many other potential forms of government investment, it makes smart grid a favorable investment product. Furthermore, smart grid deployment positively impacted employment and labor income throughout the US economy [DOE, 2013]. The reliability of smart grid is quite important, since the prevention of faults causing outages cost US$ 25 billion to US$ 180 billion annually to the US economy [Kabalci, 2016].

Other technology that has become more and more part of our everyday lives is data communication. Since the dawn of the Internet in 1960's our society has been making great leaps in technological development. Without a doubt, the Internet has accelerated progresses in various fields by interconnecting people and data around the globe, which has made sharing knowledge that much easier. In the modern society, we increasingly rely on the Internet as it has become incorporated to the way we work and enjoy our leisure time. The exchange of data has been used in a vast majority of fields and applications.

From a similar perspective, communications systems are a key component of smart grids [Arnold, 2011; Erol-Kantarci and Mouftah, 2015; Gungor et al., 2010; Gungor and Lambert, 2006; Laverty et al., 2010], being a key enabler of smart grid itself [Hauser C. H., 2005; NIST 1.0, 2010]. In [Emmanuel and Rayudu, 2016], a summary is given, based on other publications in the area, on the analogies between the Internet network and the smart grid network, shown in Table 1.1. Same as the electricity flow, the flow

Table 1.1: Analogy between the internet and evolving smart grid. [Emmanuel and Rayudu, 2016]

| Network characteristic | Internet Network | Smart grid network |
|---|---|---|
| Content | Data (IP packet) | Data and power |
| Network type | Heterogeneous and complex | Heterogeneous and complex |
| System architecture | Hierarchical, distributed and decentralized | Hierarchical, distributed and decentralized |
| Transmission flow type | Bidirectional flow of data | Bidirectional flow of data and power |
| Protocol | E.g., TCP/IP, HTTP, SMTP. | E.g., Modbus, DNP3, IP, DNP3 over TCP/IP. |
| Vulnerability/security risk | Cyber and physical | Cyber and physical |
| QoS requirements | Transparency and accessibility | Priority, critical and Latency |
| Network philosophy | Ubiquitous, adaptive and self-healing | Ubiquitous, adaptive and self-healing |
| Description | Network of networks | Network of networks |
| Services range | Multi-service | Single service |
| Storage cost | Low | High |
| Resource management | Distributed | Distributed |
| Scalability | Rapid and cheaper | Slow and expensive |
| Standards | RFCs by IETF | E.g., IEC/ANSI/IEEE |

of communication in smart grids is bi-directional and needs to meet certain specifications depending on a given smart grid application.

The communication patterns of smart grid applications are quite different from that of browsing or streaming applications that are widely used in other parts of our lives. As a result, the Quality of Service (QoS) requirements of smart grid applications differ from these traditional Internet applications, and so should the communication networks themselves. Yet, the communication networks themselves have not been structured to meet those needs.

As a result, novel networking architecture and QoS provisioning are two of the research challenges in the sector. This is because there exist various reasons for use of different technologies, depending on necessities in terms of reliability, QoS and capacity [Sooriyabandara and Ekanayake, 2010]. As communication technologies are getting more and more integrated in smart grid, reliability of communication and its impact of smart grid applications has been a relevant research direction as well [Niyato et al., 2011, 2013]. QoS of communications used in various conditions of such applications needs to be studied. First, to help decide on networking technologies to use in smart grid applications. Second, in case of wireless technologies, the deployment of base stations (BSs) needs to be studied, both in terms of position, and number of units in a specific area.

It is stated in [Fan et al., 2013] that smart grid traffic needs to reiterate its traditional research topics, namely resource allocation, routing and QoS. In [Taranetz and Rupp, 2012] for instance, future data traffic forecasts and power outage scenarios are made to define requirements of communication network capacity. And [Erol-Kantarci and Mouftah, 2015] reports, that optimal energy efficient network design in a smart grid environment has not yet been explored as well.

However, a gap is observed in the development of methodologies for optimal planning of data network for smart grid, considering the constrains of the different kinds of smart grid applications. An interesting work is presented in [Xing et al., 2016], where the authors applied graph theory to model the problem of optimal location of power-line communication (PLC) access points (APs) to connect end nodes with a service provider through PLC. Since this is an NP-hard problem, the author employed a genetic algorithm to get feasible solutions. The main constrain observed in the paper is sum of communication delays. However, in this work, the objective is to reduce the total accumulated delay and the QoS requirement to each one of the subscribers is not observed.

In a similar direction, this thesis presents a problem formulation, using graph theory, to model the search problem of optimal topology of data network. The delay constrains for each network subscriber are considered. The term QoS can generally reference various factors of communication. The rest of this work uses the term QoS to denote specifically the delay component of QoS.

## 1.2 Objectives

The above section has identified that a smart grid data network needs to be projected appropriately, so that the QoS necessities of smart grid applications are attended. On that ground, the main objective of this work is the development of a methodology, that combines clustering techniques to define the topology of a data network, with a model based on graph theory, to calculate and analyze if the QoS requirements of each network user have been met. The methodology will model a non-mobile data communication network, applicable to smart grid networks, and reflect the effect of installed networking equipment on the QoS of subscribers in a given scenario. To construct the methodology, the effectiveness of clustering techniques for this problem will first be studied.

Using sample scenarios and a software implementation of the methodology, case studies will be performed. Given the resulting QoS from the case study, an analysis will be carried out to validate the suitability of the used communication network configuration for the given scenario.

Specifically, this thesis has the following primary objectives:

- Study the communication patterns and requirements in smart grid systems;
- Perform various tests using a methodology that uses one type of network equipment, with the objective of comparing the performance and suitability of clustering techniques for network cell deployment. Next, to indicate which clustering algorithm is the most appropriate in this application;
- Develop a methodology that uses two types of network equipment: access points and gateways and uses a clustering technique suggested by the analysis from the previous objective;
- Perform a QoS analysis of both models in the form of a scenario-based case study.

The hypothesis given in this work is that it is possible to analyze capabilities of wireless network solutions for non-mobile communication networks, such as those of smart grid, in terms of delay-based QoS. Such calculations could then be performed in various configurations of the network to observe the behavior of the total QoS when the cofigurations change.

## 1.3 Work Organization

This thesis is structured into 7 chapters. After this introductory chapter follow two chapters dedicated to introduction and definition of concepts related to and used in this work, and one chapter discussing related publications. The subsequent chapters define the proposed mathematical methodology, outline specific configuration of the confronted problem and analyze the results. Finally, the last chapter concludes the work and proposes possible future work in the area. The following paragraphs offer a more detailed view of the aforementioned chapters.

Chapter 2 presents the global concept of smart grid. General definition is presented in this chapter, together with specific details about communication, standardization efforts and devices present in the network.

Clustering and specific clustering methods related to this research are defined in chapter 3. First, however, a broader view of a related subfield of Artificial Intelligence (AI) is given, to help the reader understand the role of clustering.

In chapter 4, the work related to the objectives and focus of this thesis is discussed. Previous research that this thesis expands on is presented as well.

Chapter 5 formulates the specific proposed methodology and the approach towards solving the outlined problem, and those related to it. The structure, along with the key parts of the methodology, is described in this chapter as well. This methodology uses clustering technique to plan the topology of the data network and a graph model to calculate the communication delay of each subscriber.

Chapter 6 is dedicated to studying the performance of the proposed methodology, where the clustering technique is applied in two different ways. First, a study of four clustering algorithms is performed, to assess the algorithm that is the best fit for the proposed methodology. One layer of network equipment is considered for this study. Individual positions of the equipment for every scenario is determined by a clustering algorithm. Second, in every scenario, two layers of network equipment, that relay the communication, are considered. Equipment positions in each layer are obtained by the most suited algorithm, identified in the first step, i.e., the clustering algorithm is applied twice per scenario. The performance analysis is carried out using a suite of use cases, also outlined in the chapter. Both models are evaluated, using a variety of configurations over three different smart grid scenarios, concluded by an overall discussion to summarize the model performance.

Finally, chapter 7 concludes the work by a summary of the work, followed by contributions of this work in various directions. Future research opportunities are outlined, based on the contributions of this work, where some of the topics are possible improvements of the proposed model, while others include entirely new models or approaches.

# Chapter 2

# Smart Grid

## 2.1 Initial Considerations

This chapter introduces key concepts of smart grid, especially those that are related to this thesis. Starting with motivation and a general proposal in this section, the rest of the chapter is mostly dedicated to parts of smart grid associated to communication.

The challenges of today's electricity grid, also called *power grid*, include generation diversification, optimal deployment of expensive assets, demand response, energy conservation. Further, reduction of the industry's overall carbon footprint is a great issue, as major electric energy source are fossil fuels, which leads to high greenhouse gas emissions. Moreover, only one-third of fuel energy is converted into electricity, without recovering the waste heat [Farhangi, 2010; DOE, 2018; WWF, 2010; Erol-Kantarci and Mouftah, 2015].

With the aforementioned challenges and the increased usage and society's dependency on electricity, as well as its increasing price, the evolution of the grid is necessary. The power grid system is undergoing a global innovation process towards a modern, efficient and environment-friendly, next-generation grid, called *smart grid*. This modernization is one of the most outstanding technological milestones and opportunities of the 21st century, similar to the internet evolution [Fang et al., 2012]. NIST, outlines the potential benefits and requirements of the smart grid in [NIST 3.0, 2014] using data from [DOE, 2012]:

- Enabling informed participation by customers;
- Accommodating all generation and storage options;
- Enabling new products, services, and markets;
- Providing power quality for the range of needs;
- Optimizing asset utilization and operating efficiency;
- Operating resiliently to disturbances, attacks, and natural disasters.

## 2.2 Smart Grid Communication

Smart grid uses two-way flows of electricity and information to create an automated and distributed advanced energy delivery network [Fang et al., 2012]. Because of this, many applications are being upgraded and developed to help manage the changing system. Smart grid application requirements, in terms of QoS and the ability of the existing communication technologies to meet those requirements, have been discussed in [Yan et al., 2013; Gungor et al., 2013]. In a report for United States Department of Energy (DOE), [NETL, 2008] defines four cardinal landmarks, which provide a guide to the realization of the smart grid. They include:

- Advanced Metering Infrastructure (AMI), which empowers and motivates the customer and establishes commmunications to the loads, i.e. the devices connected an energy source;
- Advanced Distribution Operation (ADO), which enables self-healing;
- Advanced Transmission Operation (ATO), which addresses congestion;
- Advanced Asset Management (AAM), which supports asset optimization for improved efficiency.

At the same time, [Fang et al., 2012] uses a different terminology and divides the entire smart grid three major subsystems:

- Smart infrastructure system,
- Smart management system,
- Smart protection system.

AMI is a key achievement of smart grid that other applications rely on. In short, an AMI system is used to measure, acquire, and analyze the data about energy consumption and power quality of each consumer [Kabalci, 2016]. It is a two-way communication network comprised of a number of integrated technologies and applications: smart meters, sensors, wide-area communications infrastructure, home (local) area networks (HANs), Meter Data Management Systems, and Operational Gateways [NETL, 2008; Mahmood et al., 2015].

In case of applications responsible for control and measurement, sensors and actuators throughout the network exchange information with the server. The most important uses of this communication are [Bakken et al., 2011]: (i) state estimation; (ii) wide-area system protection; (iii) wide-area situational awareness; and (iv) post-event analysis.

Many of these applications involve end-to-end data exchange between energy providers, energy consumers, and field devices [Sooriyabandara and Ekanayake, 2010], including sensors, actuators, energy meters, control, phasor measurement units, energy storages, Intelligent electronic devices (IEDs). IEDs, for instance, are part of last mile communication, each one installed at site of an electric feeder, further refered to as feeder.

## 2.3 Smart Grid Standards

The communication constraints of smart grid applications differ from those of a traditional Internet use. It is therefore critical that the standardization of smart grid communication takes place so that the smart grid can become a reality [Gungor et al., 2011a]. A brief comparison between the existing energy grid and smart grid is outlined in Table 2.1.

There is no widely accepted standardization and, as an effect, the development speed, integration and interoperability of smart grid applications and the involved hard-

Table 2.1: A Brief Comparison between the Existing Grid and the Smart Grid [Farhangi, 2010], [Fang et al., 2012]

| Existing Grid | Smart Grid |
|---|---|
| Electromechanical | Digital |
| One-way communication | Two-way communication |
| Centralized generation | Distributed generation |
| Few sensors | Sensors throughout |
| Manual monitoring | Self-monitoring |
| Manual restoration | Self-healing |
| Failures and blackouts | Adaptive and islanding |
| Limited control | Pervasive control |
| Few customer choices | Many customer choices |

ware has been suffering [Gungor et al., 2011b]. Moreover, the focus of the smart grid differs with region. Fortunately, many regional standardizations have been developed [Silva, 2014].

In Europe for instance, according to [ETP, 2010], the future electricity networks must be:

- Flexible: fulfilling customers' needs whilst responding to the changes and challenges ahead;
- Accessible: granting connection access to all network users, particularly for renewable power sources and high efficiency local generation with zero or low carbon emissions;
- Reliable: assuring and improving security and quality of supply, consistent with the demands of the digital age with resilience to hazards and uncertainties;
- Economic: providing best value through innovation, efficient energy management and 'level playing field' competition and regulation.

In comparison, the United States' NIST updated the definition of the key roadmap items for smart grid by the third release of the standards in [NIST 3.0, 2014]. The standards defines a number of key terms, that appear throughout the roadmap, such as Interoperability; Interchangeability; Reliability; Requirement; Resiliency; and Standards. NIST states in the standard that smart grids will ultimately hundreds of standards. For that reason, the standard prioritizes definition of cybersecurity, network communications, and seven key functionalities critical to ongoing and near-term deployments of smart grid technologies and services: (i) Demand response and consumer energy efficiency; (ii) Wide-area situational awareness; (iii) Distributed Energy Resources; (iv) Energy Storage; (v) Electric transportation; (vi) AMI; and (vii) Distribution grid management.

Seamless interoperability, robust information security, increased safety of new products and systems, compact set of protocols and communication exchange are some of the objectives that can be achieved with smart grid standardization efforts [Gungor et al., 2011a]. Smart grid activities in Research and Development (R&D) and Design and Development (D&D) are driven by researchers, government grants, and regulatory organs around the globe. According to [Silva, 2014] and [Gungor et al., 2011b], some of the most recognized standard development organizations are:

- NIST;
- DOE;
- European Technology Platform (ETP);
- UK Department of Energy and Climate Change (DECC);
- American National Standards Institute (ANSI);
- International Electro technical Commission (IEC);
- Institute of Electrical and Electronics Engineers (IEEE);
- International Organization for Standardization (ISO);
- International Telecommunication Union (ITU);
- Third Generation Partnership Project (3GPP);
- Korean Agency for Technology and Standards (KATS);
- Joint Information Systems Committee (JISC);
- CEN-CENELEC-ETSI Sector Forum on Smart and Sustainable Cities and Communities (SF-SSCC).

Table 2.2 gives a brief overview of smart grid standards, all of which are briefly discussed in [Gungor et al., 2011b]. Finally, [Fang et al., 2012; Fan et al., 2013; Chatzimisios et al., 2013] also survey several major smart grid standardization roadmaps and studies in different countries.

## 2.4 Device Standardization

Supervisory Control and Data Acquisition (SCADA) is a traditional system built for measuring substation current and voltage using remote terminal units, and current and voltage transformers [Budka et al., 2014]. With the advent of smart grid, these devices are being replaced by IEDs, discussed below. Another common type of device used in smart grid is smart meter, a two-way communication device that measures energy consumption at the appliances (electricity, gas, water or heat) [Fan et al., 2013].

### 2.4.1 IEC 61850

In [IEC 61850, 2003] NIST defined the IEC 61850, which is one of the key standards in achieving interoperability in the smart grid environment [NIST 1.0, 2010] and the standard has since been accepted in the smart grid framework worldwide for common information exchange among IEDs within power substations [Guo et al., 2016]. The standard itself was released by International Electrotechnical Commission (IEC) in [IEC 61850, 2003] and defines communication parameters of various electric substation control and monitoring applications.

IEC 61850 communications are mostly based on client / server communication, though, methods for fast, real time information exchange between IEDs like protection equipment are supported by the standard as well [Brunner, 2008]. In the network, the exchanged unit is a Piece of Information for Communication (PICOM), specified by International Council on Large Electric Systems (CIGRE) in [CIGRE WG34.03, 2001]. Each PICOM comprises of: data and description of its structure, performance requirements, and path information (logical source and destination).

Table 2.2: Overview of Smart Grid Standards [Gungor et al., 2011b]

| Type/Name of Standards | Details | Application |
|---|---|---|
| IEC 61970 and IEC 61969 | Providing Common Information Model: IEC 61970 works in the transmission domain and IEC 61969 works in the distribution domain | Energy management systems |
| IEC 61850 | Flexible, future proofing, open standard, communication between devices in transmission, distribution and substation automation systems | Substation Automation |
| IEC 60870-6 /TASE.2 | Data exchange between utility control centers, utilities, power pools, regional control centers | Inter-control center communications |
| IEC 62351 Parts 1-8 | Defining cyber security for the communication protocols | Information Security Systems |
| IEEE P2030 | A Guide for smart grid inter-operability of energy technology and IT operation with the electric power system | Customer-side applications |
| IEEE P1901 | High speed power line communications | In-home multimedia, utility and smart grid applications |
| ITU-T G.9955 and G.9956 | ITU-T G.9955 and G.9956 contain the physical layer specification and the data link layer specification | Distribution Automation, AMI |
| OpenADR | Dynamic pricing, Demand Response | Price Responsive and Load Control |
| BACnet | Scalable system communications at customer side | Building automation |
| HomePlug | Powerline technology to connect the smart appliances to HAN | HAN |
| HomePlug Green PHY | Specification developed as a low power, cost-optimized power line networking specification standard for smart grid applications | HAN |
| U-SNAP | Providing manny communication protocols to connect HAN devices to smart meters | HAN |
| ISA 100.11a | Open standard for wireless systems | Industrial Automation |
| SAE J2293 | Standard for the electrical energy transfer from electric utility to Electric Vehicles | Electric Vehicle Supply Equipment |
| ANSI C12.22 | Data network communications are supported and C12.19 tables are transported | AMI |
| ANSI C12.18 | Data structures transportation via the infrared optical port | AMI |
| ANSI C12.19 | Flexible metering model for common data structures and industry "vocabulary" for meter data communications | AMI |
| Z-Wave | Alternative solution to ZigBee that handles the interference with 802.11/b/g | HAN |
| M-Bus | European standard and providing the requirements for remotely reading all kinds of utility meters | AMI |
| PRIME | Open, global standard for multi-vendor interoperability | AMI |
| G3-PLC | Providing interoperability, cyber security, and robustness | AMI |
| SAE J2836 | Supporting use cases for Plug-in Electric Vehicles communication | Electric Vehicle |
| SAE J2847 | Supports communication messages between Plug-in Electric Vehicles and grid components | Electric Vehicle |

### 2.4.2 Intelligent Electronic Device

According to the IEC 61850 standard [IEC 61850, 2003], the IEDs refer to microprocessor-based controllers of power system equipment [Gellings, 2011]. An IED can provide multiple protective and control functions previously performed by a group of devices, like remote terminal unit or current transformer. This reduces the complexity of device topology in the substation [Gellings, 2011; Budka et al., 2014]. In 2011, IEDs using IEC 61850 were beginning to be deployed by utilities [Liu et al., 2011] and it is estimated that well over 80 % of substations will be using them by 2030 [Gellings, 2011].

IEDs are part of last mile communication and can be placed close to sensors or actuators to be monitored [IEEE 1815, 2012], or at the distribution feeder [Gong and Guzmán, 2011]. To communicate with the server, today's IEDs can take advantage of the already installed leading SCADA protocol suites, which include the traditional Distributed Network Protocol 3.0 (DNP3) and the more recently installed IEC 61850 standard, among others [McGranaghan et al., 2008; Curtis, 2005; Emmanuel and Rayudu, 2016].

### 2.4.3 Distribution Feeder

Apart from smart metering communication, another important part of smart grid is the energy distribution system itself. One of its key components is a distribution feeder or just feeder. Modern feeders can be equipped with IEDs especially to provide protection and to report the quality of energy in the system. An example diagram of three-phase feeder is shown in Fig. 2.1. Feeder is an important element, because it guides the development of the data network for smart grids, including applications and data network topology.

Radial distribution feeder, for instance, is a feeder that has only one path for the power to flow from the source – the distribution substation – to each customer. A typical distribution system consists of one or more distribution substations, each composed by one or more feeders [Kersting, 2012].

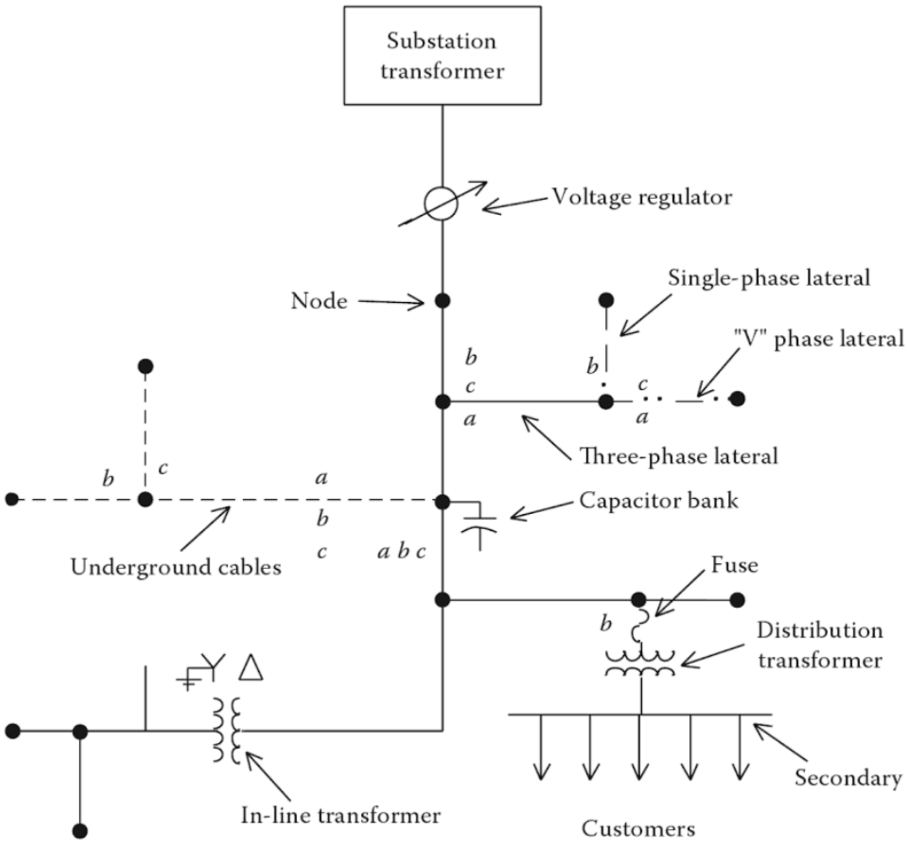



Figure 2.1: Sample distribution feeder. [Kersting, 2012]

## 2.5 Final Considerations

It can noted from this chapter, that many of the smart grid defining pieces strongly depend on standards organizations. These are often region-specific and therefore some of the definitions can vary with location. As observed, communication plays an important role in the realization of smart grid. Another important topic, related to smart grid communication is security. Thanks to the interconnectedness of the nodes using information and communication technologies, security assurance needs to be put in place. The solutions may be based on traditional internet security approaches, although penetration of the systems might often have greater negative impacts than in the case of regular internet security breaches.

# Chapter 3

# Clustering

## 3.1 Initial Considerations

Machine Learning (ML) is a field of AI, which itself is a subfield of Computer Science. It is a concept of making computers learn based on provided data, called training data. ML program is meant to study and learn from the observed data. It creates some mathematical model of knowledge, which is after used to extract some information about the observed data. The popularity and importance of ML in a variety of sectors has been growing, especially in the past decade. ML comprises of Supervised Learning, Unsupervised Learning and Reinforcement Learning. Before proceeding to the topic of clustering, this chapter also briefly introduces Supervised and Unsupervised Learning in general.

### 3.1.1 The Role of Data

Both Supervised and Unsupervised Learning are data-driven, meaning they learn based on the observed data using some algorithm. The result is a decision model, defined by the data. This is the main differentiator from traditional programming approach, where models are usually defined by a set of rules, created by humans. The goal of ML is for the model to find rules automatically based on the provided data.

It is important for data to be 'good' in various aspects, because the final model can only get as good as the underlying training data. Good data is well representative of the domain, range, as well as the noise in the data, is unbiased, does not contain irrelevant data or duplicates, and does not contain errors, such as typos, to name some of the most important ones. For this reason, raw data usually needs to be 'cleaned' and preprocessed before it is used for learning.

### 3.1.2 Supervised Learning

In the case of Supervised Learning, one is able to observe the results of the training data, i.e. the data is labeled. A Supervised Learning algorithm creates a model to approximate the real function $Y = f(X) + \epsilon$ that generated the labels $Y$ of the input data samples $X$ with some error $\epsilon$. This process is also called fitting a function $f$. Supervised Learning is divided into two categories, classification and regression.

For the process of learning, data is separated into: testing data and training data, which is often further split between pure training data and validation data. The model learns only based on the training data set, while the levels of obtained knowledge and performance of the final trained model are analyzed using the testing data, which is excluded from the learning process.

#### 3.1.2.1 Classification

Classification is the task of categorizing data into discrete groups. One of the two types of classification is binary classification, which performs a type of detection. In binary classification the ML model is learning about one specific feature and to detect whether or not it is present in the data example. Common tasks of this type include spam detection in e-mails, disease detection from medical data, fraud detection, or quality control. Some of the many classifiers, i.e. algorithms that perform classification, include neural networks, Support Vector Machines (SVM), naïve Bayes classifier and decision trees. A two-class classification example is visualized in fig. 3.1.

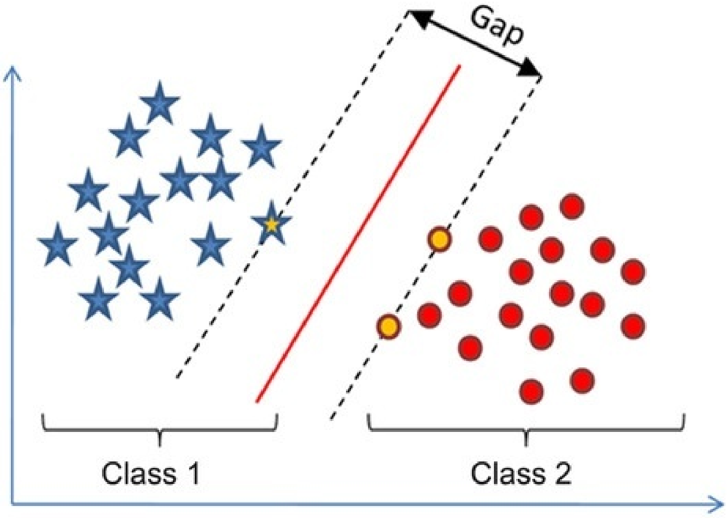



Figure 3.1: An example of classifying a data set into two classes using Support Vector Machines algorithm. Solid red line in the middle represents the separation hyperplane, i.e. the final classificator for the current problem.

Other classification types belong to the class of multi-class classification, where a set of possible outcomes is predefined and the ML algorithm aims to choose one of the possible outcomes or classes. This applies for applications like image recognition, where one might be interested in detecting the presence of more than one class in an image. An example of an Artificial Neural Network for multi-class classification is presented in fig. 3.2.

In both of the above classification types a softmax function is often applied on the results of data classification. Softmax transforms a set $m$ of $C$ arbitrary real value measurements $m_i$ into probabilities $\sigma(m)_i$ for each of the $C$ classes as

$$\sigma(m)_i = \frac{e^{m_i}}{\sum_{c=1}^{C} e^{m_c}} \, . \tag{3.1}$$

The class detection probability of $i$-th class for a sample with original measurements in vector $m$ can then be expressed by the probability $\sigma(m)_i$, where

$$\sum_{i=1}^{C} \sigma(m)_i = 1 \, , \quad \sigma(m)_i \in [0, 1] \, , \tag{3.2}$$

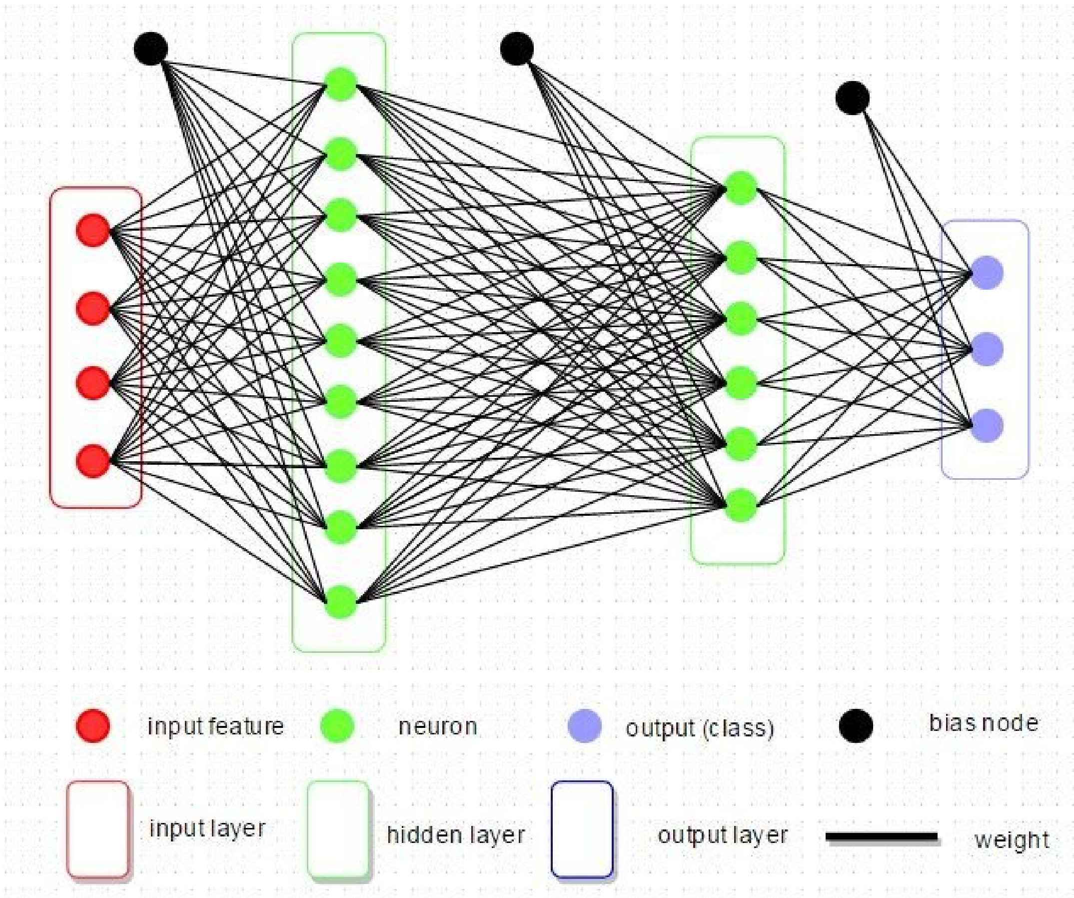



Figure 3.2: Example of an Artificial Neural Network structure for a three-class classification problem. Four nodes in the input layer mean there are four features or dimensions taken as input for every datapoint.

#### 3.1.2.2 Regression

Regression analysis, also known as regression, is a statistical technique for investigating and modeling the relationship between variables. The goal of regression is to find a regression model that describes the relationship of the variables, i.e. predictor or regressor variables, with the respective outputs, i.e. response variable $y$. Regression is often a part of a greater data analysis and provides insight on the system generating the data. Its applications occur in almost every field and common algorithms include linear regression, neural networks, nonlinear regression, or decision trees. An example of linear regression is shown in fig. 3.3.

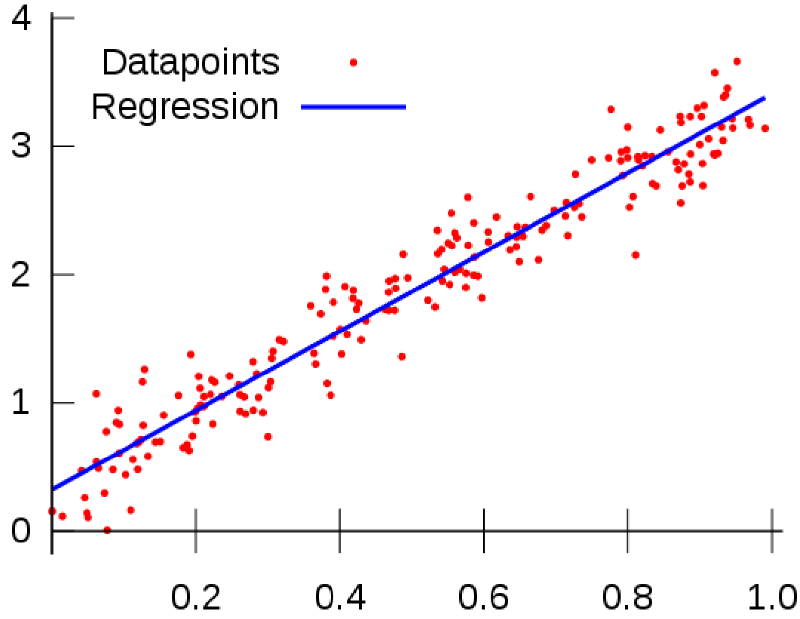



Figure 3.3: Example of regression analysis performed over a set of datapoints.

A linear regression model can map the response variable $y$ to $k$ variables $x_1, x_2, \ldots, x_k$ in a multiple linear regression model, so that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \,, \tag{3.3}$$

where $\beta_0$ is the intercept, $\beta_i, i > 0$ is the slope for the $i$-th variable, and $\epsilon$ is the error that is to be minimized. Note that the linear regression model is linear in parameters $\beta_i$, not in variables $x_i$. It is also important to note that a regression model alone, however strong empirical relationship it may show, does not imply a cause-and-effect relationship between variables [Montgomery et al., 2012], studied with the help of the Pearson coefficient, for example.

#### 3.1.2.3 Overfitting

The approximation of a function $f$, that describes the model, needs to generalize well enough to perform consistently for a variety of inputs. If the function $f$ learns to be overly precise for results of training data, it is overfitted and will not generalize well on other than training data. This means, that it may lack the ability to generalize the obtained knowledge for previously unseen data. An overfitted model yields very good results on training data, but its performance suffers when applied to testing data. When training error is much lower than testing error, the model is likely to be overfitted.

The problem of outliers is closely related to overfitting. Outliers are data samples that show abnormal or unexpected behavior (such as the value of its class label), with respect to the rest of the data. It is expected for the model to either take special measures regarding outliers or have problems returning correct results for them. Generally, such errors are allowed to occur for the purpose of combating overfitting. If a model learns to return correct results even for existing outliers, the model is likely to be overfitted.

It is desired to find a model that fits the function $f$ very well, but not necessarily perfectly, applying the concept of Occam's Razor. The concept suggests that the model should not aim to be overly precise if it would mean a significant complexity increase.

### 3.1.3 Unsupervised Learning

Information can also be extracted from data that is not paired with results, i.e. from unlabeled data. This can happen both when mapping data to correct results is impossible, and when the correct results are previously unknown. This means that no measurement of the model accuracy can be performed. It also means that instead of separating testing from training data, all data is used for learning.

The goal of Unsupervised Learning is often to find a structure or relationships present in the data. The most common type of Unsupervised Learning is cluster analysis, described in the next section. Second most common type is in the application of dimensionality reduction, which aims to reduce the data complexity by reducing the number of dimensions, while retaining as much information as possible. Two major representatives in this field are Principal Component Analysis, known as PCA and Singular Value Decomposition, known as SVD.

## 3.2 Clustering

The main part of Unsupervised Learning is clustering, also known as clusterization or cluster analysis. Its objective is to separate data into clusters. The desired final model maximizes similarity for data within clusters and minimizes similarity for data in different clusters.

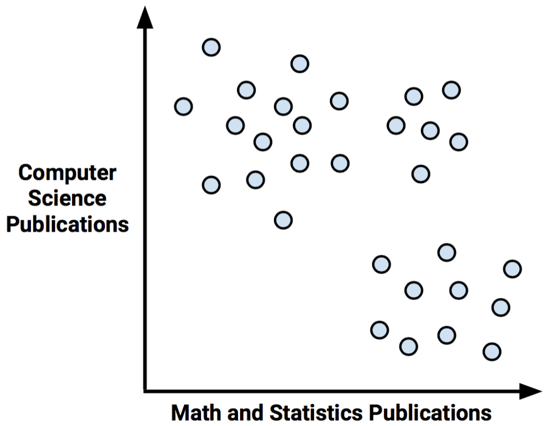
The goal of clustering is to group similar data together, while separating dissimilar ones by partitioning into clusters (groups, subsets or categories). That is to create clusters considering the internal homogeneity and the external separation. Clustering highly depends on the distance measure used [Xu and Wunsch, 2005; Everitt et al., 2001; Pavan and Pelillo, 2003]. In the case of most clustering algorithms, the resulting clusters are represented by cluster heads (CHs), also called prototypes. An example of clustering is shown in fig. 3.4.

There are many clustering algorithms and their own variations for specific problem types. These algorithms belong to clustering model types, such as centroid models, connectivity models, neural models, or density models.

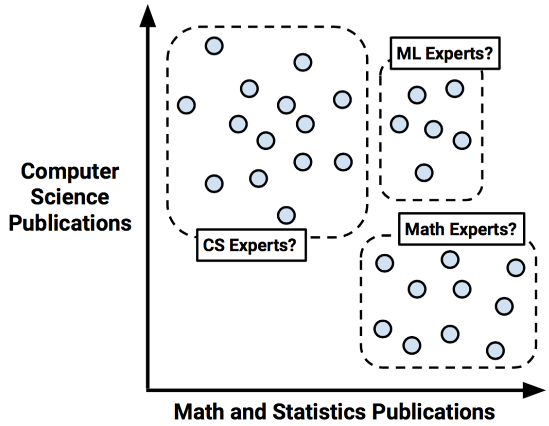
Applications of clustering include object recognition in computer vision, image segmentation and in bioinformatics, for example, sequence analysis. The main clustering methods used in this thesis are presented below. Therefore, let $n$, $k$ and $\tau$ denote the number of samples, number of clusters and number of algorithm iterations, respectively.

### 3.2.1 K-means

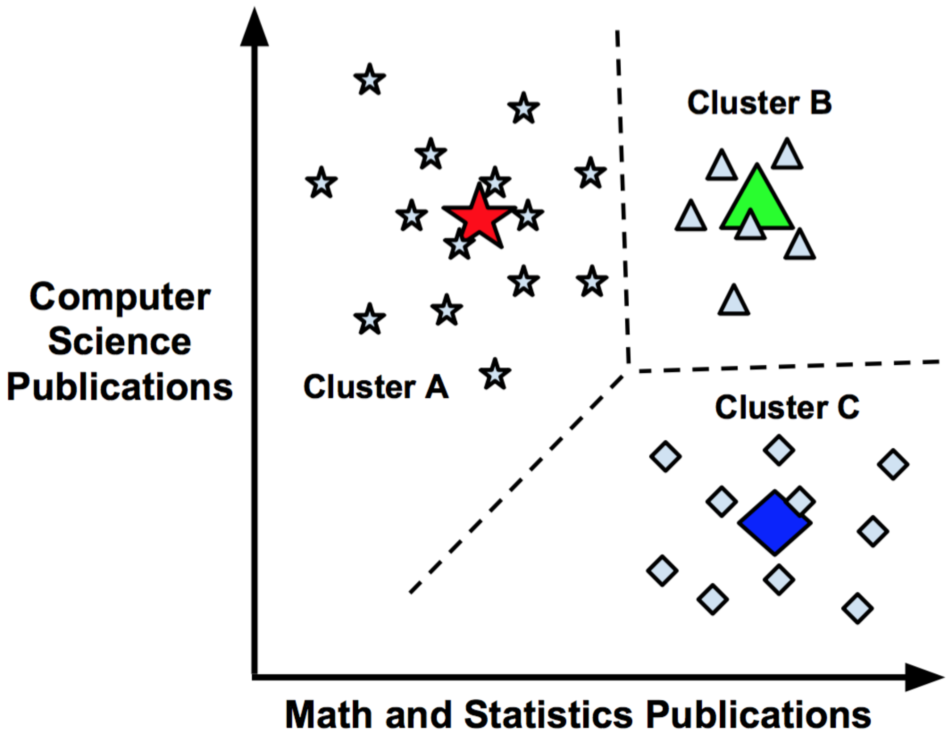
Introduced in 1957 and also known as the Lloyd's algorithm, k-means is a basic example of centroid models. It assigns a single prototype vector to each cluster. The position of the CH minimizes the sum of squared Euclidean distances between data and their cluster's CH by setting the vector in the mean of its cluster. The number of clusters $k$ is passed as a parameter and each CH is a mean of its cluster, hence the name k-means. Algorithm 1, shows how k-means calculates for optimal configurations of the $k$ prototypes.




(a) Set of two-dimensional data samples used as input.




(b) Hypothetical division of the data set into 3 clusters.




(c) Example of final positions of 3 cluster heads and the separation into clusters chosen by a clustering algorithm.

Figure 3.4: Clustering example.

**Algorithm 1** K-means algorithm

**Input**:
  $X$ - the set of observations
  $K$ - the desired number of cluster heads
**Output**:
  $\{\mathrm{CH}_k\}_{k=1}^{K}$ - the set of cluster heads
  $\{C_k\}_{k=1}^{K}$ - the clustering (partitioning) of the data, where each data point is assigned to exactly one $C_i$
**procedure** K_MEANS($X, K$)
  1. Initialize the cluster heads
  **for** $k \leftarrow 1$ **to** $K$ **do**
    $\mathrm{CH}_k \leftarrow$ initialize randomly
  **repeat**
    **for** $k \leftarrow 1$ **to** $K$ **do** ▷ assign to closest cluster
      $C_k \leftarrow \{x \in X : \forall j, \|x - \mathrm{CH}_k\|^2 \leq \|x - \mathrm{CH}_j\|^2\}$
    **for** $k \leftarrow 1$ **to** $K$ **do** ▷ update the cluster heads
      **if** $|C_k| > 0$ **then**
        $\mathrm{CH}_k \leftarrow \frac{1}{|C_k|} \sum\limits_{x \in C_k} x$
      **else**
        $\mathrm{CH}_k \leftarrow$ re-initialize
  **until** $\forall k : C_k^{\tau+1} = C_k^{\tau}$ ▷ convergence reached

Common technique used with k-means is re-initialization. It increases the probability of finding global minima by choosing different initial CH values. However, it cannot be guaranteed that the algorithm finds the global optimum. As to initialization, one of the strategies to set the initial CH vectors is to draw the vectors from the set of observations without repeating the draws. The time complexity of k-means is relatively low at $O(nK\tau)$. Its disadvantage is the inability to handle categorical data and the fact that the algorithm is not suited for non-convex cluster shapes.

### 3.2.2 K-medoids

Another example of centroid models, k-medoids, also known as Partitioning Around Medoids (PAM), was introduced in 1987. It works similarly to k-means and is shown in Algorithm 2. However, instead of assigning a mean vector to each of the $k$ clusters, it assigns its median value. This means finding a member of each cluster that is closest to the cluster mean, thus minimizing the absolute error [Aggarwal and Reddy, 2016]. The value of the integer $k$ is a required input parameter for the algorithm.

Compared to k-means, k-medoids is more robust by being less sensitive to outliers and extreme values. This is compensated by the algorithm's rather high time complexity of $O(K(n-K)^2)$.

**Algorithm 2** K-medoids algorithm

**Input**:
    $X$ - the set of observations
    $K$ - the desired number of cluster heads
**Output**:
    $\{\mathrm{CH}_k\}_{k=1}^{K}$ - the set of cluster heads
    $\{C_k\}_{k=1}^{K}$ - the clustering (partitioning) of the data, where each data point is assigned to exactly one $C_i$
**procedure** K_MEDOIDS$(X, K)$
    1. Initialize the cluster heads
    **for** $k \leftarrow 1$ **to** $K$ **do**
        $\mathrm{CH}_k \leftarrow$ initialize randomly
    2. Iterate
    **repeat**
        **for** $k \leftarrow 1$ **to** $K$ **do** ▷ assign to closest cluster
            $C_k \leftarrow \{x \in X : \forall j, \|x - \mathrm{CH}_k\|^2 \leq \|x - \mathrm{CH}_j\|^2\}$
        Randomly select a nonrepresentative example $x_i$.
        Compute the total cost $S$ of swapping the representative example $m$ with $x_i$.
        If $S < 0$, then swap $m$ with $x_i$ to form the new set of representative examples.
    **until** $\forall k : C_k^{\tau+1} = C_k^{\tau}$ ▷ convergence reached

### 3.2.3 Agglomerative Hierarchical Clustering

There are two types of hierarchical clustering: agglomerative and divisive [Jain and Dubes, 1988]. Divisive opts for a top-down approach, where initially all observations are in one common cluster. The algorithm then selects a cluster for division, divides it into two, and repeats these steps until every observation is in its own cluster.

The Agglomerative Hierarchical Clustering (AHC), shown in Algorithm 3, chooses a bottom-up approach. AHC starts with every data point being in its own cluster. The algorithm then merges two clusters that are somehow "closest" to each other. The algorithm continues merging while updating a dissimilarity matrix until all the observations are in the only cluster remaining [Aggarwal and Reddy, 2016].

There are various linkage types one can use when dividing or merging, i.e. linking, the clusters and in fact any valid distance measure can be used. The most common ones for this application include minimum, maximum, mean or weighted mean distance between the elements of the clusters, and Ward's method. Ward's method minimizes the inner squared distance of the clusters being merged, resulting in the lowest increase of variance upon merging.

AHC and other hierarchical algorithms often represent the resulting cluster structure as a dendrogram (a tree of clusters). It demonstrates the division or merge performed at each level of the algorithm and shows the resulting clusters for a dendrogram cut at any level. In the case of AHC, cutting the graph at level $l$ results in $n - l$ clusters and the algorithm can also be stopped once the desired level is reached. The time complexity of AHC is at least $O(n^2)$.

**Algorithm 3** Agglomerative Hierarchical Clustering

```
Input:
    X - the set of observations
Output:
    Dendrogram representing clusters for each k ∈ {1...n}.
procedure AHC(X)
    1. Compute the dissimilarities between all examples from X and
          store them in the dissimilarity matrix
    2. Iterate
    repeat
        Merge clusters as C_{a∪b} = C_a ∪ C_b
        Insert a new row and column containing the distances
            between the new cluster C_{a∪b} and the remaining clusters.
    until One maximal cluster remains.
```

### 3.2.4 Affinity Propagation

The last clustering algorithm presented in this thesis is *affinity propagation*, shown in Algorithm 4. Similarly to k-medoids, the CHs used by this algorithm are some of the data samples themselves. In contrast, unlike the algorithms presented above, affinity propagation does not receive the integer $k$ as a parameter. Although the value of $k$ can be influenced by the diagonal of the given similarity matrix, the number of clusters is decided by the algorithm [Cheng and Day, 2014].

**Algorithm 4** Affinity Propagation

```
Input:
    X - the set of observations
    s - the similarity matrix storing at s_{i,j} how similar is x_i to x_j, ∀i ≠ j ∈ {1...n},
        while s_{i,i} stores the preference of x_i becoming an exemplar ∀i ∈ {1...n}.
Output:
    {CH_k}_{k=1}^K - the set of cluster heads
    {C_k}_{k=1}^K - the clustering (partitioning) of the data, where each data point
        is assigned to exactly one C_i
procedure AffinityPropagation(X)
    1. Initialize availabilities a(i,k) to zero ∀i,k
    2. Iterate
    repeat
        Update all the responsibilities given the availabilities
        Update all the availabilities given the responsibilities
        Combine availabilities and responsibilities to obtain the exemplar decisions
    until Termination criterion is met.
```

Before the algorithm is executed, the numerical preferences of each observation becoming a prototype are chosen and passed as an argument to the algorithm. Setting all preferences to the median of all data points is often applied in practice.

The algorithm itself has a time complexity of $O(n^2\tau)$ and can be thought as about as passing messages between pairs of data points. These messages express the suitability of one sample to be the exemplar of the other. The message values are being iteratively updated until the communication converges to a final set of clustering prototypes.

There are two types of messages that are passed during the algorithm. The first is the responsibility $r(i,j)$, sent from $i$ to $j$, which is the accumulated evidence that sample $j$ should be the exemplar for sample $i$. The second one acts as a response to message $r$ and is sent from $j$ to $i$. It is the availability $a(i,j)$ and is the accumulated evidence that sample $i$ should choose sample $j$ to be its exemplar. This message considers the values for all other samples that $j$ should be an exemplar. This way, the exemplars in the resulting set of CHs are both similar to many samples and chosen by many as representative of themselves [Frey and Dueck, 2007].

## 3.3 Final Considerations

This chapter is focused primarily on clustering, which is a technique this work relies on. To put clustering in the context of ML, other ML cathegories were briefly presented in this work as well. For more information about AI and ML, refer to [Russell et al., 2003] and [Mitchell, 1997]. In comparison to other ML techniques, clustering is similar to classification in the sense that its goal is to separate the data into groups. In clustering, however, there is no label available for the datapoints, or, as in the case of this thesis, there is no need for labels. The result or accuracy of clustering can be difficult to measure. In some cases, one measures the clustering result based on some external function that uses the clustering output, as is the case of this thesis. Alterations of clustering parameters, such as the number of clusters, are often made to reach global optimum over the possible configurations.

# Chapter 4

# Related Work

## 4.1 Initial Considerations

In the recent years, two interconnected works were published that have the objective of optimizing QoS in a data network in smart grids. Both of them use PLC as a communication technology and a standardized scenario with 13 users to study the results. The types of the equipment and their position are also equal.

First, [Silva, 2014] studies the effect of the network topology on the communication QoS in a smart grid scenario using an analytical model. The network has pre-defined communication equipment and their available positions. A search is performed in the space of topologies using a genetic algorithm, where every chromosome represents one topology. The best of the found topologies satisfies the QoS requirements of 10 out of 12 users. The author uses a restriction of communication, which considers only a subset of messages to be transferred, thus lowering the network traffic. In this case, the network reaches a total QoS of 100 %. Two more analitical models are also introduced, that are used to perform an optimization, using a Markov Decision Process, to determine the optimal policy for controlling the use of a shared data rate.

In [Júlio, 2015], the state space of possible topologies was searched using a genetic algorithm and Dinitz's max flow algorithm was used to calculate the maximum flow in the network graph. The obtained results show that in most cases, a suboptimal topology is found, resulting in 2 out of 12 users with exceeded time delays. Another topology with 100 % is also found; however, with less frequency. The work also compares the result to those using a brute-force approach, also analyzing the execution time. The author suggests that the work can be utilized by future research by using the same graph model and substituting other transmission technologies, such as mesh Wi-Fi, Worldwide Interoperability for Microwave Access (WiMAX) and optical networks.

In a similar direction, this thesis presents a problem formulation using graph theory to model the search problem of optimal topology of data network. This thesis also shares the objective of minimizing the total QoS of the network in a given scenario. To compare the works better, one of the scenarios used in this thesis, denoted as Scenario $A$, is equal to the one both of these publications used.

## 4.2 Related Work

This section discusses recent publications in the area of smart grid network deployment strategies or multi-level communication deployment, which are related to this thesis, as well as research challenges in smart grid communications.

In [Heo and Varshney, 2003], the authors propose an intelligent energy-performance algorithm for cluster-based wireless sensor network (WSN), the Intelligent Deployment and Clustering Algorithm (IDCA). The work compares the results to a random deployment and to Distributed Self-Spreading Algorithm, which uses a peer-to-peer mechanism. IDCA combines cluster structuring and the same peer-to-peer deployment scheme. The main measures of performance are coverage, uniformity, and time and distance of convergence. The paper uses mobile nodes to cover a rectangular area. Nodes know their own position and decide where to move next. Further, nodes only perform peer-to-peer communication with nodes in their direct communication range. The algorithm balances between performance increase and preservation of energy. IDCA uses local clustering, a concept introduced in [Kawadia and Kumar, 2003] and [Lin and Gerla, 1997], in regions of density that is close to that of uniform distribution. In local clustering mode, nodes with high amount of remaining energy in their neighborhood are moved. If the local density is too high or too low, the algorithm chooses a peer-to-peer mode and updates the location of a node based on the calculated partial force. The results show that IDCA with the local clustering approach successfully reached uniform distribution in the area and total area coverage. In all other metrics (convergence distance and its standard deviation, time, and uniformity) IDCA performs similarly or even outperforms the pure peer-to-peer approach. Importantly, the average distance travelled by a node is reduced by more than a factor of two, while the standard deviation is also lower. This means lower energy use as well as a longer network lifetime.

Cellular network design method using Simulated Annealing (SA) algorithm is presented in [Ekici and Ersoy, 2001]. The research introduces the use of SA as an optimization algorithm to determine multi-tier network that minimizes the implementation cost of the network. The study focuses on cellular network for mobile users, where techniques, such as *handoff*, i.e. switching connection from one cell to another, need to be taken into consideration. Therefore, the effect of user mobility, call duration, and call arrival rates are some of the principal parameters of the study. SA is compared with Greedy Search (GS) algorithm and a random solution generator, Generate and Test (GAT), as a solution reference. Results show that in most cases of the scenario configuration the SA algorithm outperforms the GS algorithm. Comparing the solution quality with GAT shows that SA is able to find a solution either better than GAT or performing very similarly. Finally, the proposed directions of future work include other optimization algorithms, other neighborhood functions for the SA algorithm, and studying more than two levels of the network.

A downlink heterogeneous multi-tier cellular network is modeled in [Dhillon et al., 2011] to study the coverage probabilities in a region. Importantly, the research considers the presence of multiple types of BSs in the network, whether it is macro-, micro-, pico- or femtocell. Signal-to-interference-plus-noise ratio (SINR) distribution calculation is used to obtain the coverage and outage probabilities in any location. The nodes are mobile and connect to the BS with the highest SINR, selecting from all BS types. One of the main points the work investigates is the influence of interference on coverage. The space is fully tessellated following a maximum SINR connectivity model, resulting in a coverage plot resembling a circular Dirichlet tessellation. The model accuracy is verified through comparisons with an actual 4G macrocell deployment, concluding that the proposed deployment model is as good as the hexagonal grid model.

One of the engineering and research challenges in the communications technology area of smart grid in [Sooriyabandara and Ekanayake, 2010] is QoS provisioning. According to the article, state of the art QoS concepts considered by most communications networks and application developers will not be sufficient in the smart grid networks. Guaranteed delivery and strict deadlines on message-level delivery times are some of the restrictions that will have to be met. The article adds, that this raises need for defining and developing solutions considering application, service and infrastructure requirements of the smart grid. Techniques to integrate and implement these novel QoS requirements will also need to be developed.

Observing the recent publications in the area, much research is being conducted to: study the impacts of control and communication system vulnerabilities on power systems under contingencies [Rahnamay-Naeini et al., 2012]; evaluate, through prototyping or modeling, the feasibility of using several technologies for transmitting data for the purpose of implementing smart grids, e.g. [Noorwali et al., 2016; Müller et al., 2012; Kheaksong et al., 2016]; characterize QoS requirements of smart grid applications, e.g. [Bakken et al., 2011; Alcatel-Lucent, 2010]; propose conceptual architectures for the implementation of data transmission networks capable of meeting the restrictions of the smart grid applications, e.g. [Vallejo et al., 2012; Islam and Lee, 2012; Asbery et al., 2016]; develop algorithms and methodologies to ensure compliance of new protocols, like Internet Protocol version 6 (IPv6), with the QoS requirements for these different classes of applications, e.g. [Chakraborty and Chaki, 2015].

## 4.3 Final Considerations

Considering the analized publications in this chapter, it is observed that smart grid networks require detecting an optimal topology. The search must consider communication QoS metrics and is to be performed in the planning stage of the network deployment.

The publications identified as related to this topic are few and far between. This can be seen in the previous section, where few publications have been identified that relate to the topic. More importantly, no work has been found to study or improve the planning strategy with regards to QoS in smart grids, with or without using computer science techniques to automate the optimization process. This indicates the lack of academic research in the area, which corresponds to and supports the identified research gaps in the same area.

It is therefore considered that the proposal of this thesis is original and contributes with a network deployment strategy model that studies QoS in the network. The outcome of this work can be used by utility companies, looking to deploy a non-mobile network, such as that in smart grid.

# Chapter 5

# Proposed Methodology

## 5.1 Initial Considerations

The smart grid components, including the specific communications, are developed and specified by the research community around the world. It is necessary to ensure that QoS requirements are reached by applications of the smart grid defined by standards, such as IEC 61850. Therefore, in this chapter, the methodology proposed in this thesis to analyze the problem of planning of an adequate data network topology is presented. This methodology combines a clustering technique to define the data network topology, and graph theory to compute the maximum delay for each subscriber. The proposed approach computes transmission delays per subscriber based on the subscriber specifications and the network topology. In the rest of this thesis, the terms user, subscriber and IED are used interchangeably.

## 5.2 Motivation and Objective

The objective of this research is to develop a practical methodology, which is used to analyze the QoS of the messages involved in the communication of smart grid applications. The methodology is developed to perform the following steps, given a scenario defined according to section 5.3:

1. Determine positions of $k_1$ access points and $k_2$ gateways using a clustering algorithm, where numbers $k_1$ and $k_2$ are given. Do so based on the positions of the network subscribers;

2. Define optimal topology for upstream and downstream communication;

3. Calculate the data rate on each communication link in the network, based on the topology from step 2.

4. Given the amount of data to transfer per network subscriber, calculate delay for each subscriber and direction of communication. Compare this delay to a given maximum delay.

5. Evaluate if QoS was reached for each subscriber and display results.

Given the results from step 5, an analysis needs to be carried out to validate the suitability of the used communication network configuration for the given scenario.

In this work, QoS is the key performance indicator of the network suitability in a given scenario. Using explicit amounts of data to transfer, paired with maximum delay, the methodology determines whether or not are the messages of a subscriber delivered within the predefined time limit.

## 5.3 Scenario Definition

The communication in the proposed methodology is divided into four layers, as is visualized in Figure 5.1. The IEDs, that are specified in the methodology input by coordinates, make up the lowest layer, $\text{layer}_0$. Two wireless network tiers continue, with access points that form the following layer, $\text{layer}_1$. Next, $\text{layer}_2$ is composed by gateways. Finally, the last layer, $\text{layer}_3$, is what represents the control center that manages this network, including the distribution feeders. Also, let $\text{hop}_1$, $\text{hop}_2$, and $\text{hop}_3$ denote the communication between $\text{layer}_0$ and $\text{layer}_1$, $\text{layer}_1$ and $\text{layer}_2$, and $\text{layer}_2$ and $\text{layer}_3$, respectively.

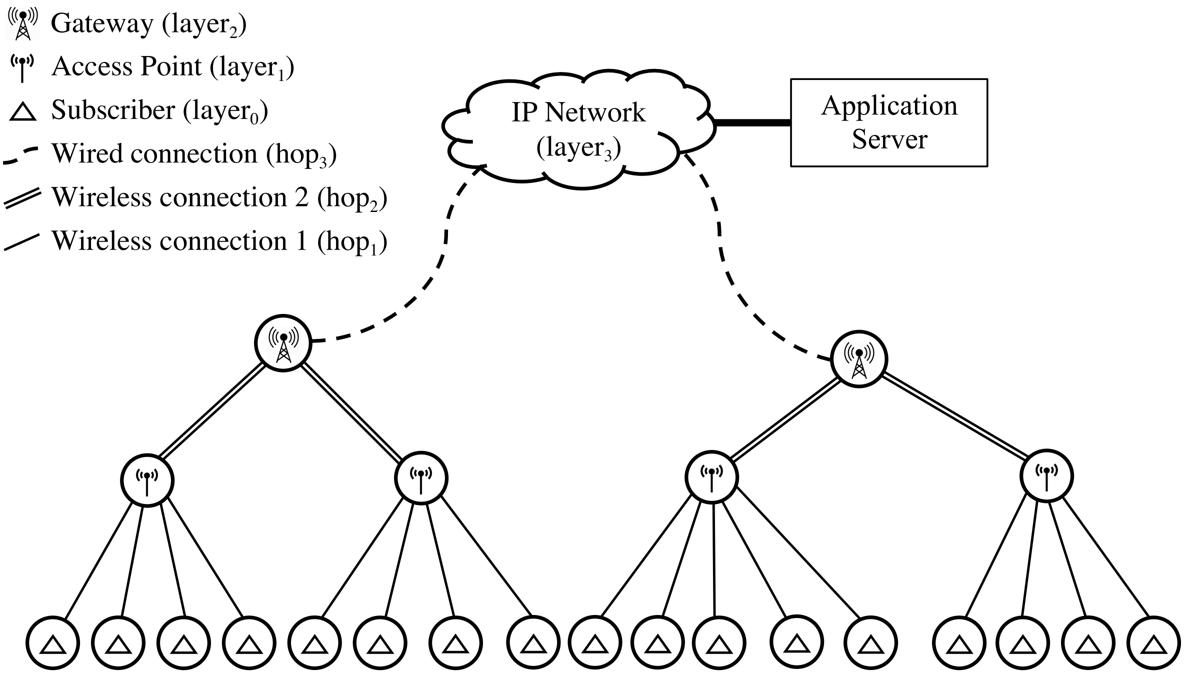



Figure 5.1: Communication structure of the proposed methodology.

### 5.3.1 Subscriber Configuration

Two-dimensional Cartesian coordinates define the key part of the scenarios. Together with the coordinates, each node is specified by the size of its messages and maximum tolerated delay for both uplink and downlink. Nodes can be also split into groups and these properties would be defined per group.

Rather than separately for each message, the QoS is calculated per network subscriber. This is to express the importance of delay requirements satisfaction of all of the user's messages in order to reach positive QoS. Moreover, previous research [Silva, 2014; Júlio, 2015] uses this type of grouping as well. Therefore, for both uplink and downlink each user type has a defined amount of bits to transfer, which is a sum of its message sizes in the direction, while the maximum tolerated delay per user is the minimum of its defined message delays in that direction. One user communication type needs to be assigned to every subscriber. In a real scenario analysis, the relation between communication types and subscribers must to be done according to real applications and subscribers. In this thesis, it is performed using roulette-style probabilities (probabilities with a total sum of 1).

To summarize, the methodology uses the following subscriber configurations to define a scenario:

1. Two-dimensional Cartesian coordinates of network subscribers;
2. Communication types that specify the amount of data to transfer and maximum tolerated delay, separately for uplink and downlink;
3. Roulette-style probabilities used to assign one communication type for each subscriber.

### 5.3.2 Network Configuration

The network proprieties are determined by specifying the number of access points and gateways, and describing the communication technologies used in each communication hop.

Wireless technologies rely on detailed specification to calculate the channel capacities, which includes the operating frequency, bandwidth, number of frequency bands, and power and gain per equipment category in every hop.

Finally, the methodology also requires that Time Division Duplex (TDD) is specified by the fraction of time allocated for uplink and downlink, represented by $\text{TDD}_{\text{up}}$ and $\text{TDD}_{\text{down}}$ respectively, where $\text{TDD}_{\text{up}}, \text{TDD}_{\text{down}} \in [0, 1]$. The methodology puts no restriction on the sum of $\text{TDD}_{\text{up}}$ and $\text{TDD}_{\text{down}}$, although in most use cases $\text{TDD}_{\text{up}} + \text{TDD}_{\text{down}} = 1$.

To summarize, the methodology uses the following network configurations to define a scenario:

1. Clustering algorithm to use;
2. $k_1$ and $k_2$, i.e. the number of access points and gateways respectively;
3. Wireless communication technology for $\text{hop}_1$ and $\text{hop}_2$: operating frequency, bandwidth, number of frequency bands, number of TDD time frames;
4. Power and gain of wireless equipment in $\text{layer}_0$, $\text{layer}_1$ and $\text{layer}_2$;

5. Data rate in $hop_3$;

6. TDD ratio used in all hops.

## 5.4 Methodology Components

### 5.4.1 Clustering

The set of coordinates serves as an input of $n$ data points for a clustering algorithm, which is performed twice. First, the $n$ coordinates are clustered using a specified algorithm, creating $k_1$ CHs, whose coordinates form $layer_1$. Second, the $layer_1$ set serves as an input for the second round of clustering. This generates a set of $k_2$ CHs, that represent $layer_2$. Sets $layer_1$ and $layer_2$ determine the positions of access points and gateways respectively.

Clustering algorithms used in the proposed methodology perform clustering purely based on the provided coordinates. That means the CHs are also defined by two-dimensional coordinates. These coordinates are then used as network cell positions. If the used clustering algorithm is based on Euclidean distances, this distance is then an actual Euclidean distance, i.e. distance in a straight line, between two points.

It is important to study which clustering algorithms are the most suitable for this application. This is because algorithms vary and so do the values of CHs. Given that all clustering algorithms are able to process two-dimensional set of examples, most clustering algorithms can be used for this methodology.

For the sake of optimal network discovery, it is preferred that the number of clusters is user-defined, although it is not mandatory. It is obligatory, though, that the used algorithm assigns exactly one cluster to each data point. Fuzzy algorithms, which use partial cluster memberships, are not allowed.

### 5.4.2 Graph Topology and Delay Calculation

Every two nodes that transfer data directly between each other rely on a communication channel formed between them. Channel capacity describes the qualities of a connection between two nodes. In wireless communications, it is a result of various characteristics of the specific connection, such as distance, present obstacles and other active radio equipment, weather conditions, and the installed radio equipment of the two nodes.

In the following steps, the delay calculation of a message is outlined. In this section, let *up* and *down* denote upstream and downstream respectively. First, the data rates per network link, represented by edges of $G_{\mathrm{up}}$ and $G_{\mathrm{down}}$, need to be calculated. After, a network load is modeled, in which all nodes communicate simultaneously. Upstream and downstream communication is calculated and modeled separately. Uplink and downlink can share the same channel and the methodology can be adapted accordingly.

To analyze delays on a communication link, data rates that describe the amount of bits per second, need to be calculated. One of the main building blocks of this calculation is SINR. Specifically, SINR of the receiving node is used in this thesis. For the calculation of SINR, refer to appendix A. With that, the data rates of the links are calculated, in the case of wireless communication links. The bitrate is calculated separately for every edge in both graphs $G_{\mathrm{up}}$ and $G_{\mathrm{down}}$.

Data rate is the capacity of a graph edge, given by a calculation of maximum channel capacity of the link, using the Shannon-Hartley equation. The value of the maximum capacity of a channel $C_{i,j}$ from node $i$ to node $j$ is:

$$C_{i,j}(\mathrm{bps}) = \beta_{i,j} \log_2(1 + \mathrm{SINR}_{i,j}) \,, \tag{5.1}$$

where $\beta_i$ is the bandwidth between node $i$ and $j$ in Hz, and $\mathrm{SINR}_{i,j}$ is the SINR of the channel from $i$ to $j$ as perceived by $j$.

By applying the calculation procedure, the obtained bitrate of $C$ is in the units of bits per second, or bps for short. Given the bitrate, it needs to be taken into account that the network links are shared communication channels and that the delay of a message per link is calculated by considering all of the present traffic in the channel. This applies for wired technologies as well, and the only exception are links that form $\text{hop}_1$, where the channel is indeed always used only by one subscriber. The delay $t_{i,j}$ of any message on an edge $i$ in a direction $j$ is given as

$$t_{i,j} = \frac{f_{i,j}}{C_{i,j}\mathrm{TDD}_{\mathrm{dir}}}\,, \tag{5.2}$$

where $f_{i,j}$ stands for the total amount of traffic flowing through the edge in the direction from $i$ to $j$ and $\mathrm{TDD}_{\mathrm{dir}}$, $\mathrm{dir} \in \{\mathrm{up}, \mathrm{down}\}$ is the ratio of time units, such as time slots, assigned for the direction dir, where up and down stand for uplink and downlink respectively. The total delay $t^{\mathrm{P}}_{i,j}$ on path from $i$ to $j$ in either $G_{\mathrm{up}}$ or $G_{\mathrm{down}}$ equals to the sum of the delays on its edges:

$$t^{\mathrm{P}}_{i,j} = \sum_{t' \in \Psi_{i,j}} t'\,, \tag{5.3}$$

where $\Psi_{i,j}$ is the set of delays on the path from $i$ to $j$.

### 5.4.3 Results and Visualization

Exposing the calculated results is a crucial part of the proposed solution. A well-informative result formulation is important, so that the usefulness of the methodology is not reduced.

The principal result unit is a vector of delays per user. For the purposes of this work, this vector is compared to the vector of maximum allowed delays per user to mark which users reached the QoS limit and which did not. The sum of users of below-limit QoS are then exposed in the output itself.

Using the result of the vector comparison from the previous paragraph, and all the known coordinates, a graph of all subscribers, access points and gateways is created to visualize the resulting positions, as well as to show which subscribers are below the QoS limit, and which are above it. If the methodology identified that some users have insufficient QoS, the utility company can decide to provide a different form of communication in the case of these users, following the visualization, an example of which is shown in Figure 5.2.

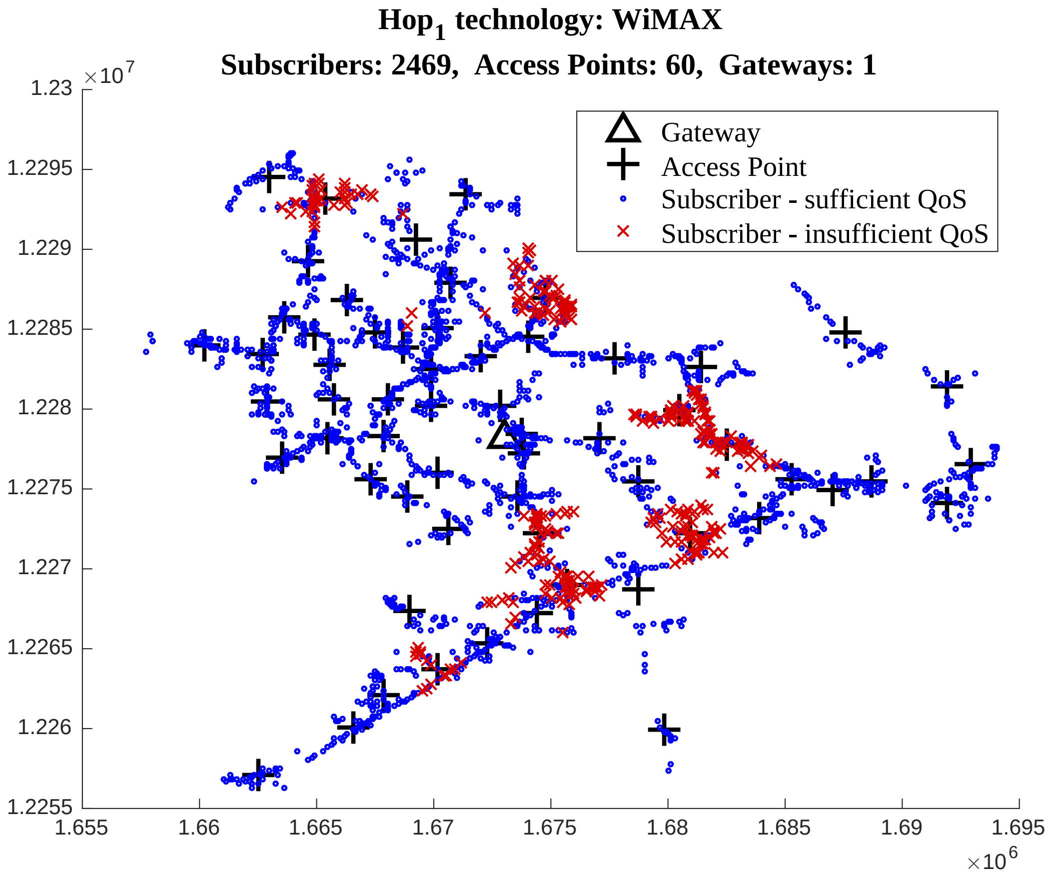



Figure 5.2: Example of a result visualization, showing access point and gateway positions as well as 2469 network subscribers and their QoS state.

Results also include the formed topology by the methodology, as well as the exact gateway and access point coordinates and the data rates on each link. The coordinates can be used as an aid to deploy an actual network in the given geographical area.

### 5.4.4 Restricted Mode

A restricted mode is also possible, yet it does not make part of the methodology itself. It is rather a technique that can be applied through input manipulation. To enable the restricted mode, the coordinates are kept the same, but the amounts of data are lowered to a desired level before passing the values as an input to the methodology procedure. In some cases, that may mean changing the maximum allowed delays as well.

The restricted mode is useful for removing less important data from the communication and letting only the high priority messages use the communication networks, thus avoiding high volumes of data traffic. This represents a critical situation when the power system is unstable and only the applications related to protection and control of the power network must be performed.

## 5.5 Final Considerations

This chapter presented a methodology for communication networks that finds an optimal topology and calculates delay per user.

Note, that the values of the methodology parameters used in this thesis can be modified. Even various parts of the methodology can eventually be substituted if necessary to adapt the methodology for different use cases. For example:

- The cell position determination is performed by a clustering algorithm. Not only can various clustering algorithms be used, but another approach can be used as well. Alternatively, the positions can also be defined before execution, which is helpful if the cells are already installed;

- Various wireless technologies can be used, given that the channel capacity is adequately calculated and TDD ratio set accordingly;

- Technologies other than wireless can be used. For wired technologies, different clustering algorithms can be studied that better represent the wireline objectives, given that the wires are not laid in a straight line, but rather in a sequence of straight line sections;

- The channel capacity calculation can be omitted altogether if desired. In this situation, the adjacency matrix, which holds the data rates on the network links can be parametrized;

- Various types of smart grid applications and other network communication types can be used. The key input pair per subscriber is the amount of data, coupled with maximum allowed delay, for both upstream and downstream.

# Chapter 6

# Case Study

## 6.1 Initial Considerations

After describing the concepts and techniques in the previous chapters, this chapter defines configuration of the model from Chapter 5, concrete used test cases, as well as the performance analysis itself. The quality measure for the case study is a delay-based QoS. Each round of testing is evaluated based on the number of IEDs that are unable to complete the data transfer with delay within the established maximum time for both uplink and downlink.

Before presenting the two-level clustering model introduced in the previous chapter, the following sections present the results of a single-level version of the model. The single-level model served for determining the optimal clustering technique to use for this application by comparing four clustering algorithms. Apart from the number of layers that the models use, they both work the same way. It is important to note, that the network equipment strengths were not device-class specific in the single-layer model. For that reason the model has been redesigned making the two-level version more accurate. Both models share configuration values from section 6.2, with each model having its specific configurations seen in their respective sections.

## 6.2 Common Configuration

### 6.2.1 Coordinates

The coordinates of scenarios used in this thesis are released at IEEE Power & Energy Society$^{TM}$ (PES) website[1] for testing purposes in disciplines that utilize electric feeders. These scenarios serve as a basis for studies, e.g. [Silva, 2014; Júlio, 2015], and are described in the following sections of this chapter.

Models in this thesis were tested using a small scenario, Scenario $A$, so that the results can be compared to [Silva, 2014; Júlio, 2015]. Two considerably more challenging scenarios, Scenario $B$ and Scenario $C$, where the exact two-dimensional coordinates are available, were also used to evaluate the proposed model.

### 6.2.2 Communication Patterns

Specific requirements such as delay, amount of data and message importance were studied to define the communication requirements. The messages are defined in a form of a PICOM. Each application then uses different message types. This in turn influences the amounts of data to send, together with the maximum allowed delays. In accordance with [Silva, 2014], three specific message classes were chosen from the IEC 61850 definition [IEC 61850, 2003]:

- **Type 1 – High-speed messages** involve sending simple binary codes to IEDs, such as: close, start, stop, block, unblock, trigger, and others. Messages related to network protection have maximum tolerated delay between 1 ms and 10 ms, while those related to automation functions have tolerated delay between 20 ms and 100 ms.
- **Type 2 – Medium-speed messages** carry data with high precision of the time of measurement, yet the maximum tolerated delay between the IED and the server can be as high as 100 ms. These messages generally carry system state information specify a time when the receiver should react.

[1] https://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html

- **Type 3 – Low-speed messages** are used to transfer system data, event logs, system thresholds, including alarms for events related to measurements of non-electrical quantities, like equipment temperature and pressure. The tolerated delay for this message class can be as high as 500 ms.

Following the above definitions, each message type was represented by one of the PICOM types. The payload sizes, maximum delays and other PICOM information are presented in Table 6.1.

Table 6.1: Details of message classes, defined using PICOMs. Adapted from [IEC 61850, 2003].

| Message Class | PICOM | Importance | Payload (bits) | Maximum Delay (ms) |
|---|---|---|---|---|
| Type 1 | 15 | High | 1 | 1 |
| Type 2 | 2 | Medium | 128 | 50 |
| Type 3 | 3 | Low | 1024 | 100 |

Also, all messages use a set of protocols: Real Time Protocol (RTP) / User Datagram Protocol (UDP) / Internet Protocol (IP), which adds a 40 byte overhead for each message. This overhead is added to every message payload, increasing the total message size substantially. Moreover, as stated in Chapter 5, each network subscriber (IED) is represented by the sum of message sizes, while the total maximum tolerated delay is the minimum tolerated delay of the messages of a given user.

Note, that the application of Protection uses type 1 messages for uplink and type 2 for downlink. It is also observed that in [Silva, 2014; Júlio, 2015] Scenario $A$ is also used and the assigned applications per IED are known. To maintain the network configuration in larger scenarios with more IEDs, the application assignments are scaled so that the above ratios are preserved. Both the ratio of these assignments among IEDs, denoted as IED use ratio, and the three user types from above are summarized in Table 6.2.

### 6.2.3 Restricted Mode

Observing the message types above, it may be beneficial for the system to decouple type 1 and type 3 messages. The reason is that while type 1 carries small amount of data

Table 6.2: Message sizes, including protocol overhead, and maximum delays per communication category.

| Application | PICOMs used | IED use ratio | Direction | Total size of messages (bits) | Maximum delay (ms) |
|---|---|---|---|---|---|
| Protection | 2, 15 | $\frac{5}{12}$ | Uplink | 321 | 1 |
| | | | Downlink | 448 | 50 |
| Quality of Energy | 3 | $\frac{4}{12}$ | Uplink | 1344 | 100 |
| | | | Downlink | 0 | — |
| Protection and Quality of Energy | 2, 3, 15 | $\frac{3}{12}$ | Uplink | 1665 | 1 |
| | | | Downlink | 448 | 50 |

in up to 1 ms, type 3 carries larger amount of data in up to 100 ms. This means when both messages need to be transmitted at the same time. The data rate on the path between such IED the server needs to be high enough to send both messages in up to 1 ms. Further, type 1, type 2, and type 3 messages are rated as high, medium, and low importance respectively. Exploring this fact, this thesis uses a restricted mode that restricts the communication in the network to only type 1 and type 2 messages, excluding the low importance messages of type 3. The type 3 messages are assumed to be sent later to decrease the network traffic. Restricted mode is therefore also part of the QoS analysis.

## 6.3 Scenarios

IEDs have the role of a network subscribers or users. Each IED is installed at site of a feeder. The feeders' positions are defined as an input by two-dimensional Cartesian coordinates and define a key part of the scenarios.

### 6.3.1 Scenario *A*

The original test feeder paper [Kersting, 2001] defines IEEE 13 Node Test Feeder, referenced in this work as Scenario *A*. It contains 12 user nodes, used IED positions in this work, and one access point node, through which the communication flow is meant to

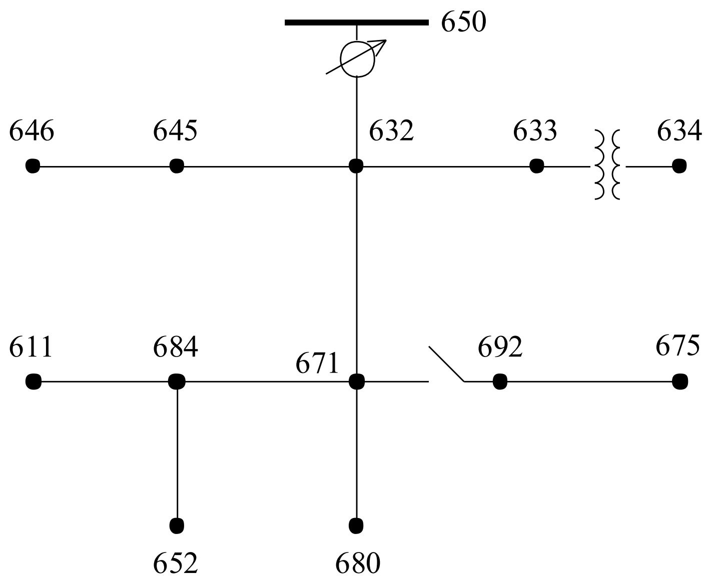



Figure 6.1: 13 feeders of Scenario $A$ [Kersting, 2001]. In [Silva, 2014] the node at the top is used as an access point and is excluded from the scenario in this thesis.

enter and exit the network. The dimensions of the scenario are 400 m × 1600 m. As is discussed in Chapter 4, the work of [Silva, 2014] and [Júlio, 2015] uses this scenario for communication optimization as well. In both of these cases, PLC is used as the networking technology.

In the case of Scenario $A$ the node coordinates are not explicitly defined. Instead, the scenario is described by distances between the individual nodes. Concrete coordinates were reconstructed using these lengths and the published scheme of the network, shown in Fig. 6.1, which was used for understanding in which direction each distance applies. During the coordinate reconstruction, it was considered that the angle between the edges is either 180° or 90° as is depicted in the scheme. The calculated coordinates themselves are presented in Table 6.3.

### 6.3.2 Scenario $B$

Scenario $B$, shown in Fig. 6.2, is the densest of the tested scenarios of this thesis, with 906 IED nodes in an area of 160 m × 150 m. On IEEE PES website, this scenario is referred to as European Low Voltage Test Feeder and represents a typical European

Table 6.3: Coordinates of 12 nodes from Scenario $A$.

| IED | x (m) | y (m) |
|---|---|---|
| 646 | 300 | 1000 |
| 645 | 147.6 | 1000 |
| 632 | 452.4 | 1000 |
| 633 | 300 | 390.4 |
| 634 | 56.16 | 1000 |
| 611 | 452.4 | 1000 |
| 684 | 208.56 | 390.4 |
| 671 | 300 | 390.4 |
| 692 | 300 | 85.6 |
| 675 | 117.12 | 390.4 |
| 652 | 208.56 | 146.56 |
| 680 | 452.4 | 390.4 |

feeder deployment. It has great demands in terms of communication. This is a challenging scenario, since the high density will also apply in the case of the installed cells.

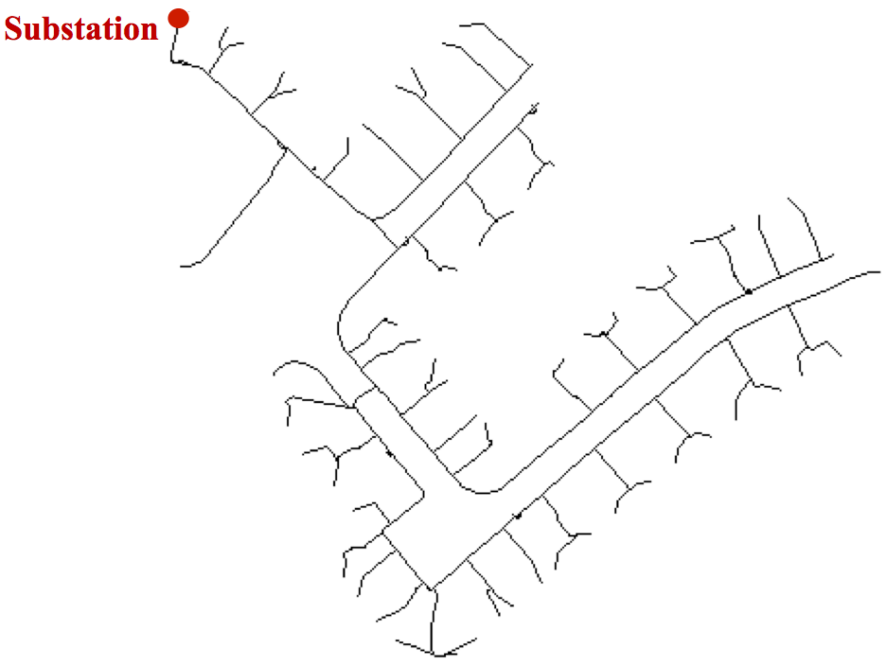



Figure 6.2: European Low Voltage Test Feeder coordinates used for Scenario $B$ [IEEE PES, 2015].

### 6.3.3 Scenario $C$

The largest scenario of the three both in size and in number of IEDs is Scenario $C$, shown in Fig. 6.3. Even though it is called 8500-Node Test Feeder scenario, it features 2469 primary (medium voltage) nodes for which the coordinates are given. The density of the nodes is not high, since the scenario dimensions are 4000 m × 4500 m.

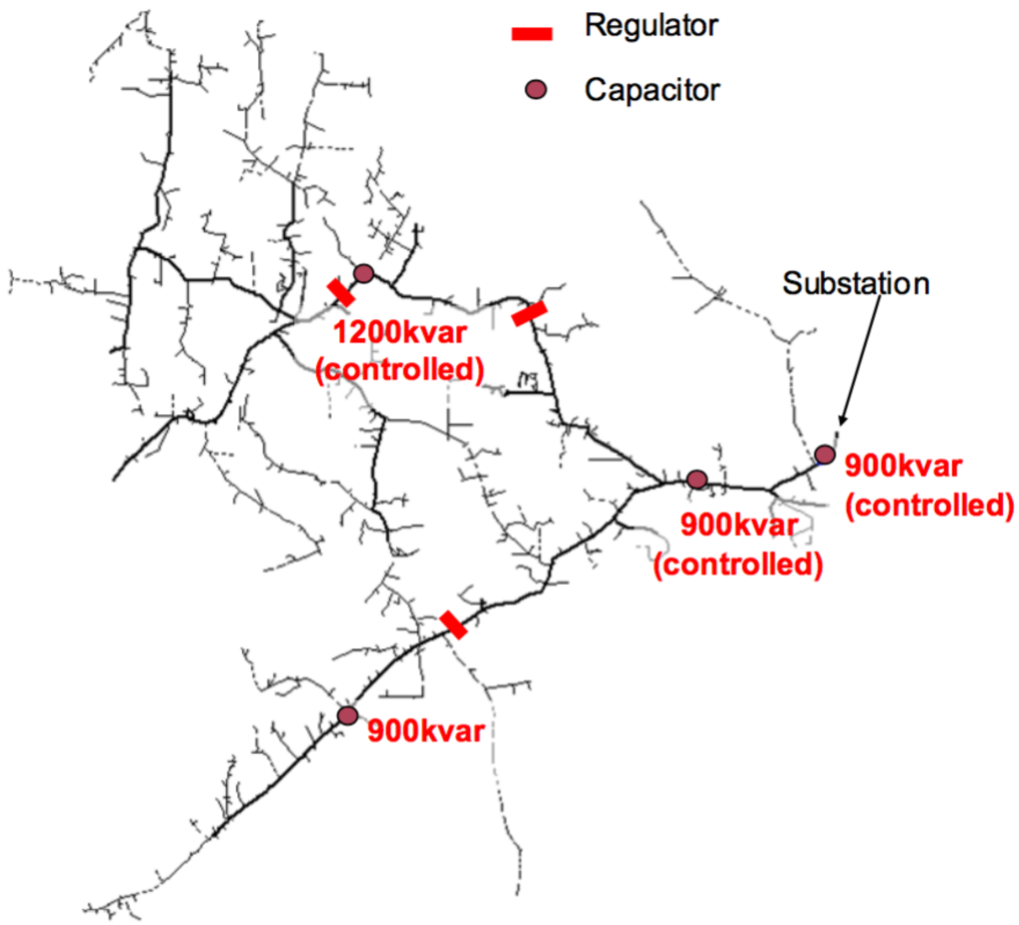



Figure 6.3: Scenario of around 8500 feeders, 2469 of which were used in Scenario $C$ [Arritt and Dugan, 2010].

## 6.4 Single-Level Model Configuration

### 6.4.1 Model Preparation

This section presents the comparison of four clustering algorithms for the application of deployment positioning of base station equipment for smart grid. The model used in this case uses one layer of BSs to attend the IED communication necessities. It is configured according to section 6.2 with some minor changes:

- **Networking** - Technology between IEDs and BSs: WiMAX; frequency: 2.5 GHz; radiated power: 43 dBm; gain: 0 dB; bandwidth: 10 MHz; technology between BSs and the system servers: wired with a fixed 1 Gbps transfer rate in both directions.
- **Coordinates** - The coordinates from Scenario $A$ and Scenario $C$ were used for this model.
- **Transfer rate calculation** - Free-space propagation model with white noise and foliage was used, where tree crown radius $R = 3\,\mathrm{m}$.

#### 6.4.1.1 Clustering Configuration

The configuration of clustering algorithms is briefly shown here. Four algorithms were used for this model: k-means, k-medoids, AHC, and affinity propagation as described in Chapter 3.

K-means is configured to set the initial cluster centers as values of randomly selected data samples. To escape local minima, the algorithm is executed 10 times, every time with different initial CHs. The clustering used as a final solution is the one that minimizes the total sum of Euclidean distances between data points and their respective clusters.

Next, k-medoids uses Euclidean distance metric for distance calculations. To select the initial $k$ samples as CHs, k-means++ algorithm, is used.

The AHC algorithm uses Euclidean distances as well. Five linkage types are used for comparison: *average*, *complete*, *single*, *Ward*, and *weighted*.

To complete the configuration, affinity propagation uses a damping factor of 0.9. Damping factor (between 0.5 and 1) is the extent of oscillation allowed when updating the values. That is, higher values lead to lower oscillations. As a similarity matrix $s$, a symmetrical matrix of negative Euclidean distances between points, was used. This matrix, with zeros on the main diagonal, is used throughout the algorithm to quickly obtain the distances between samples. The algorithm also receives a preference coefficient for every data sample, which serves as an initial indicator of the sample being selected as a CH. The coefficient for an $i$-th sample is obtained as the median value of the $i$-th column of $s$.

### 6.4.2 Results

#### 6.4.2.1 Scenario $A$

First, in Scenario $A$ with 12 IEDs, the smart grid applications (introduced in Table 6.2) per IED are based on [Silva, 2014] and [Júlio, 2015], shown in Table 6.4. These assignments of communication types to the IEDs are unchanged in this scenario.

Table 6.4: Smart grid applications assigned to the 12 nodes of Scenario $A$. Each application defines the communication patterns according to Table 6.2

| IED | Application |
|---|---|
| 646 | Protection |
| 645 | Protection and Quality of energy |
| 632 | Protection |
| 633 | Protection and Quality of energy |
| 634 | Quality of energy |
| 611 | Protection |
| 684 | Protection |
| 671 | Quality of energy |
| 692 | Protection and Quality of energy |
| 675 | Quality of energy |
| 652 | Protection |
| 680 | Quality of energy |

In this scenario, all IEDs' delay constraints were attended with using only one BS, i.e. $k = 1$ regardless of clustering technique used. For higher number of BSs, inter-cell interference of WiMAX was not high enough to lower the QoS. Affinity propagation chose to group into 3 clusters and was also able to attend all IEDs. In summary, in Scenario $A$ all algorithms performed perfectly without any difference.

#### 6.4.2.2 Scenario $C$

Scenario $C$ is characterized by a high number of IEDs. For this reason, a higher number of clusters was chosen. For the sake of generality, the IED communication categories that assign one of the three application types, as depicted in Table 6.2, are randomly reassigned 10 times for each model configuration.

K-means and k-medoids yielded the best results, attending all 2469 IEDs in all configurations for all three tested values of $k \in \{10, 20, 30\}$. In comparison, affinity propagation selected a high number of clusters, $k = 94$, which resulted in high QoS, close to 100 %. AHC was unable to meet QoS necessities of all IEDs, especially with *single* linkage type, which performed significantly worse than the other ones.

In restricted mode affinity propagation attended all IEDs. As to the AHC, all linkage types except for single either came very close to attending all or managed to attend all. This restricted mode is important for maintaining QoS of communication of high importance. The results of AHC and affinity propagation of both full communication mode and the restricted mode are presented in Table 6.5, which summarizes all of the 10 runs per each table row by mean and standard deviation values of the results.

#### 6.4.2.3 Summary of Results

In this research, clustering is performed using only two-dimensional coordinates of the IEDs as an input. The algorithms' performance was then evaluated according to the QoS levels reached by the constructed network. K-medoids and k-means led to perfect QoS in this model. K-medoids' CH positions are drawn from the set of IED positions, therefore, k-medoids can be financially more beneficial to use in this application, since a BS deployment can share resources with an already installed feeder and its respective IED, leading to lower costs of network deployment (CAPEX).

Table 6.5: QoS results of Agglomerative Hierarchical Clustering and affinity propagation clustering algorithms. Values represent units of IEDs with sufficient (Attended) and insufficient (Not Attended) QoS.

| Clustering method | Cluster count | Stream direction | Regular mode Attended (mean) | Regular mode Not Attended (mean) | Regular mode Not Attended (Std. dev.) | Restricted mode Attended (mean) | Restricted mode Not Attended (mean) | Restricted mode Not Attended (Std. dev.) |
|---|---|---|---|---|---|---|---|---|
| AHC average | 10 | up | 2300.4 | 168.6 | 22.40 | 2465.3 | 3.7 | 3.09 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC complete | | up | 2354.5 | 114.5 | 10.15 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC single | | up | 897.7 | 1571.3 | 22.22 | 2238.9 | 230.1 | 112.71 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC ward | | up | 2343.7 | 125.3 | 9.49 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC weighted | | up | 2310.2 | 158.8 | 20.73 | 2461.7 | 7.3 | 1.16 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC average | 20 | up | 2418.5 | 50.5 | 3.34 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC complete | | up | 2417.5 | 51.5 | 6.65 | 2466.7 | 2.3 | 3.02 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC single | | up | 985.7 | 1483.3 | 23.71 | 1532.5 | 936.5 | 98.10 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC ward | | up | 2398.6 | 70.4 | 17.61 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC weighted | | up | 2338.6 | 130.4 | 17.29 | 2466.8 | 2.2 | 0.92 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC average | 30 | up | 2400.0 | 69.0 | 11.56 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC complete | | up | 2394.3 | 74.7 | 13.47 | 2456.0 | 13.0 | 1.76 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC single | | up | 1012.3 | 1456.7 | 21.41 | 1159.0 | 1310.0 | 20.02 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC ward | | up | 2415.0 | 54.0 | 16.24 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| AHC weighted | | up | 2349.8 | 119.2 | 9.20 | 2467.1 | 1.9 | 1.52 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |
| Affinity Propagation | 94 | up | 2468.6 | 0.4 | 0.52 | 2469.0 | 0.0 | 0.00 |
| | | down | 2469.0 | 0.0 | 0.00 | 2469.0 | 0.0 | 0.00 |

Other clustering techniques offer acceptable results, but considering the results of k-medoids and k-means, these two are the best choices among the tested algorithms. Affinity propagation cannot be recommended because of the high number of clusters it selected and the low QoS performance.

The reason for the success of k-means and k-medoids comes from how the four algorithms approach the task of clustering. K-means and k-medoids have advantage in that their implicit optimization is the sum of intra-cluster distances and nothing more. Unlike affinity propagation and AHC, these two algorithms do not take into consideration the relationships between cluster members, between entire clusters, or some other relation, such as suitability of a data point to become a CH. This is where AHC and affinity propagation loose the lead when focusing on the aforementioned clustering features that are irrelevant for this specific application. The objectives of k-means and k-medoids algorithms indeed showed to be the best, in comparison to the other tested algorithms.

## 6.5 Two-Level Model Configuration

### 6.5.1 Communication Network

Considering $hop_1$ requires a shorter-range connection, Wi-Fi was one of the technologies chosen as for this hop. Given the fact that Wi-Fi may not always offer sufficient reach and the technology of choice for $hop_2$, WiMAX was also included for deployment in $hop_1$. Both technologies are deployed in separate model configurations to compare their performance.

In $hop_2$ WiMAX technology was used in all configurations, because of its long-range capabilities and therefore higher suitability for an inter-cell communication. The parameters of both Wi-Fi and WiMAX communication networks are shown in Table 6.6.

Given that a lower number of gateways necessary, $hop_3$ is suitable for wired connections. This connection is executed using fiber optic connection, considering a transfer rate of 10 Gbps between the gateways and the control center for both uplink and downlink.

Table 6.6: Parameters of wireless technologies and their values.

| | Wi-Fi | WiMAX |
|---|---|---|
| Frequency (GHz) | 2.4 | 3.5 |
| Bandwidth (MHz) | 20 | 10 |
| Frequency channels (count) | 11 | 11 |

### 6.5.2 Antennas

Considering there are two technologies used for $hop_1$, there are six types of network equipment used for the proposed model that are specific per network layer and the

Table 6.7: Values of wireless network equipment parameters.

| Hop | Direction | Transmitter | Technology | Power (dBm) | Gain (dB) |
|---|---|---|---|---|---|
| 1 | Uplink | IED | Wi-Fi | 20 | 0 |
| | Uplink | IED | WiMAX | 10 | 3 |
| | Downlink | Access Point | Wi-Fi | 40 | 0 |
| | Downlink | Access Point | WiMAX | 30 | 15 |
| 2 | Uplink | Access Point | WiMAX | 30 | 15 |
| | Downlink | Gateway | WiMAX | 43 | 20 |

technology used. The selected parameters specific for each level are shown in Table 6.7. It can be observed, that one of the value pairs of power and gain repeats twice in the table, owing to the fact that the WiMAX access points in the proposed model use the same antennas for upstream and downstream. Note, that it is assumed that the gain is available in all horizontal directions around the antenna. This can be accomplished for example by using a sector antenna [Kitchen, 2001].

### 6.5.3 Time Division Duplex

To share a communication link, TDD is deployed to manage the shared link. According to [3GPP, 2012], UMTS Terrestrial Radio Access - Time Divison Duplexing (UTRA-TDD) system uses 15 time slots per frame. Also, [Nasreddine and Lagrange, 2003] uses UTRA-TDD, while assigning two slots for control messages and the remaining 13 slots for the application.

However, in this work, the communication protocol overhead has already been included in the message sizes. Therefore, this thesis assigns all 15 slots to the application. It is also assumed that all 15 slots are available on every link, where each slot serves for either upstream or downstream. Each scenario includes various values of the TDD ratio. In section 6.6 the TDD ratio represents the assignment of slots to upstream and downstream, which translate to $\text{TDD}_{\text{up}}$ and $\text{TDD}_{\text{down}}$ respectively.

### 6.5.4 Delay Calculation

The calculation of channel capacity and message delays is described in Chapter 5 and values of the parameters are given in this chapter. Starting with the values of the communication patterns and networking technologies in the sections above, this section concludes with the remaining parameters and assumption statements necessary for the calculation.

First, in a calculation of inter-cell interference it is assumed that IEDs do not generate signals strong enough to cause significant interference, given the lower power of the network equipment. It is also considered, that every cluster communicates on one of the frequency bands given in Table 6.6. The specific band for each cluster is randomly chosen and it is assumed that equal bands and neighboring bands create full interference, while other cases create no interference, where neighboring bands are frequency bands directly next to each other, such as 2 and 3, but not 2 and 4.

Further, to optimize the intra-cell interference, wireless links within the same cluster are assumed to use channels orthogonal to each other, increasing the system capacity by creating no interference between one another. This can be achieved, for instance, by deploying Orthogonal frequency-division multiple access (OFDMA) modulation scheme.

Second, the value of the wired communication speed given above is not reduced by channel calculations or data management protocols. Instead, the full value is used as the actual speed of the link. The calculation of the channel capacity and transfer speeds is performed only for the wireless links.

Third, free space propagation model with foliage was used to generate noise and losses, as described in Chapter 5. Thermal noise and losses caused by foliage, atmospheric attenuation and rain attenuation were considered. Values of the parameters to calculate these variables are listed in Table 6.8.

### 6.5.5 Clustering

Considering the results above and the fact that k-medoids restricts the search space of CH vectors to the data samples themselves, k-means is the algorithm that has

Table 6.8: Parameters describing the communication environment that influences signal propagation.

| Parameter | Value |
|---|---|
| $T$ | 20 °C |
| $R$ | 3 m |
| $\alpha_{\text{atm}}$ [Rappaport et al., 2011] | 0.01 dBm km$^{-1}$ |
| $\alpha_{\text{rain}}$ [Qingling and Li, 2006] | 0.1 dBm km$^{-1}$ |

the potential to achieve the best performance among the compared algorithms. Therefore, k-means clustering algorithm is used as the only technique in the multi-level version of this model. The configuration of the k-means algorithm is the same as in the case of one-level model, given in section 6.4.1.1.

## 6.6 Two-Level Model Case Study

This section describes the configuration of each scenario together with the results obtained. For the sake of generality, the feeder categories that assign one of the three application types, as depicted in Table 6.2, are randomly reassigned 100 times for each model configuration, except for Scenario $A$, where the assignments were kept equal to [Silva, 2014] and [Júlio, 2015].

Four TDD ratios were used in each scenario: (i) 7 : 8; (ii) 10 : 5; (iii) 12 : 3; and (iv) 14 : 1. Communication technology used in hop$_1$ is Wi-Fi and WiMAX, while hop$_2$ only uses WiMAX. The number of access points and gateways, i.e. values of $k_1$ and $k_2$ respectively, varies per scenario.

### 6.6.1 Scenario $A$

Observing the results of the single-level model from section 6.4, the configuration of Scenario $A$ is simple. First one with two access points and one gateway, i.e. $k_1 = 2$, $k_2 = 1$, and second one with three access points and one gateway, i.e. $k_1 = 3$, $k_2 = 1$.

After the analysis of the single-layer model results, it was expected that the multi-level one was going to perform well in this scenario. Indeed, all tested configurations managed to transfer all data within the maximum delays. Even the access point count increase did not cause significant data rate changes, keeping the delay QoS at 100 %. As there were no differences between the results of the configurations, it can be concluded that all configurations perform well in such a low demand task. An example of network cell positions chosen by the model is shown in Fig. 6.4.

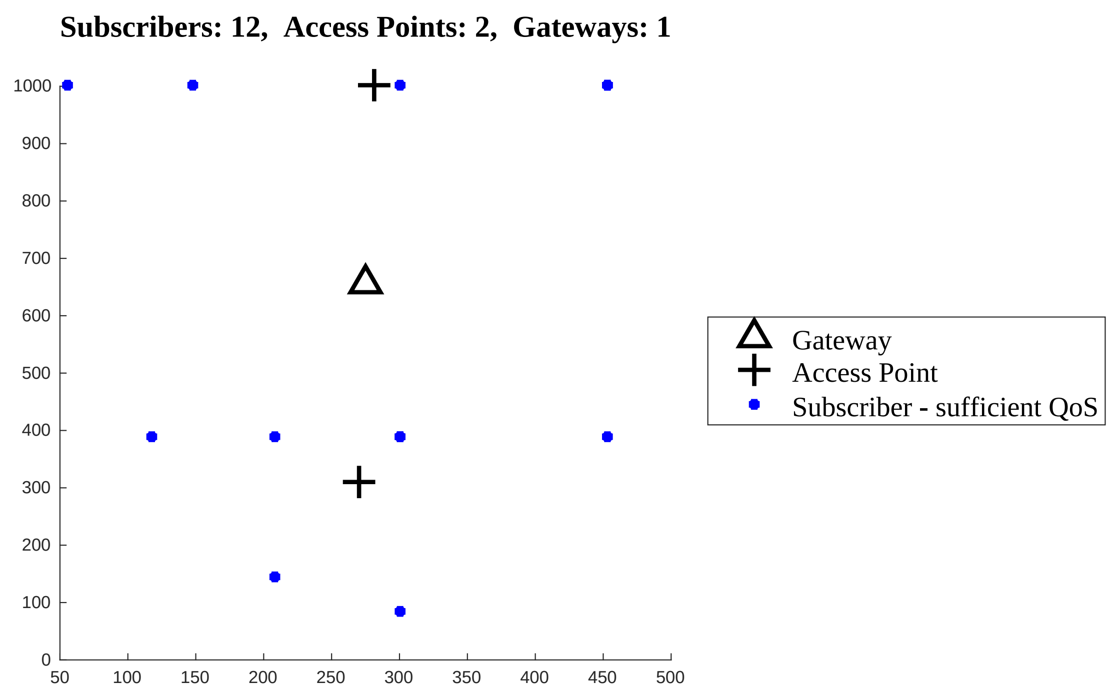



Figure 6.4: Example result of network cell positions relative to subscribers as chosen by the two-layer model for Scenario $A$

### 6.6.2 Scenario $B$

For the densest scenario in the test suite, it is interesting to see the effects of congestion and interference as the cell count changes. Tables 6.9, 6.10, 6.11, and 6.12 list the configurations used and the respective results obtained.

The results for downstream communication show that the delay constrains in this direction are satisfied in most cases. This is because of higher allowed delays of 50 ms for transferring only 448 bits of data. The analysis in the following paragraphs considers only communication in the upstream direction.

The minimum cell count with 100 % QoS, among the tested ones, is with $k_1 = 20$, $k_2 = 1$ for the regular mode, and $k_1 = 10$, $k_2 = 1$ for the restricted mode, considering a strong TDD ratio of $14 : 1$. With TDD ratio $10 : 5$ for example, Wi-Fi needs to be used and $k_2$ needs to rise to the value of 40 to reach the same performance.

Depending on how much does one need to avoid the occasional low QoS, the optimal cell count can be smaller by a good amount. For example, for TDD ratio $12 : 3$ and using Wi-Fi in $\text{hop}_1$, 10 access points get reasonably good results, with total QoS $\geq 99.79\,\%$. The resulting positions and QoS results of this particular configuration for regular and restricted mode are visualized in Fig. 6.5 and Fig. 6.6 respectively.

### 6.6.3 Scenario $C$

Given its large size, Scenario $C$ is modeled for various cell count configurations and requires more cells to be installed. Even then, fewer cells are also put into test to analyze the QoS changes with the cell count increase. The summary of the configurations used and the respective results obtained is shown in Tables 6.13, 6.14, 6.15, and 6.16.

The delay constrains of downstream communication are satisfied in most of the configurations. Same as in Scenario $B$, this is because of the higher allowed delay in this direction. The analysis in the following paragraphs focuses purely on upstream communication.

For WiMAX, the minimum cell count that yields 100 % QoS is with $k_1 = 200$, $k_2 = 1$ for the regular mode, considering TDD ratios $12 : 3$ and $14 : 1$. The $10 : 5$ TDD ratio almost managed the same, yet in one of all 100 instances of the execution of this configuration, one user does not have the QoS limit satisfied, leading to only 0.01 below-limit QoS, on average. Cluster sizes $k_1 = 40$, $k_2 = 1$ perform well enough to attend the QoS necessities of all subscribers in the restricted mode with $10 : 5$ TDD ratio. To demonstrate the effects of communication restriction, the following configuration is selected: $k_1 = 100$, $k_2 = 1$, WiMAX in $\text{hop}_1$, TDD ratio $10 : 5$. The resulting positions and QoS results of configuration for regular and restricted mode are visualized in Fig. 6.7 and Fig. 6.8 respectively.

In the case of Wi-Fi, the QoS never reached 100 % in the tested scenarios for regular mode. The restricted mode gets perfect score with $k_1 \in \{40, 60, 80\}$, $k_2 = 1$, depending on the selected TDD ratio. When using the regular mode, examples of a good total QoS is $k_1 = 100$, $k_2 = 2$ with TDD ratio 10 : 5, and $k_1 = 200$, $k_2 = 2$ with TDD ratio 12 : 3 that reach average QoS levels of 99.28 % and 99.86 % respectively. The highest QoS reached using Wi-Fi is 99.98 %, where the configuration utilizes TDD ratio 14 : 1 and $k_1 = 200$, $k_2 = 2$.

### 6.6.4 Summary of Results

Overall, it can be concluded that:

- TDD ratios have big effect on results by managing the resources assigned for uplink and downlink;
- Increasing the AP count generally improves the total QoS, while the increase of gateway quantity improves the QoS in some cases and lowers it in others, depending on the scenario;
- The use of Wi-Fi in $\text{hop}_1$ proved to be beneficial in Scenario $B$, compared to WiMAX. On the other hand, WiMAX was a better choice for Scenario $C$. This is because the WiMAX configuration uses 10 MHz bandwidth, as opposed to 20 MHz for Wi-Fi. The narrower bandwidth results into higher SINR values, which in turn increases the total data rates. This effect is strong in long-distance communication, such as the one in Scenario $C$, where it leads to better results;
- Downstream communication QoS is not a big problem in this smart grid application, as it reaches perfect QoS levels in almost all test cases. This is due to the large allowed delays and relatively low message sizes;
- The restricted mode brought the number of not attended IEDs to zero in almost all instances and is a useful tool when studying alternative communication scheduling.

## 6.7 Final Considerations

This chapter introduced the scenarios, configurations and test cases used in the execution of the proposed model. After identifying the best clustering techniques for this problem, a case study of the model, as well as a summary of the obtained results is performed.

The results from this chapter support the feasibility of using wireless technologies for smart grid networks and similar non-mobile networks in general, especially compared to the performance of PLC from [Silva, 2014; Júlio, 2015].

Table 6.9: Number of IEDs with insufficient QoS in Scenario $B$ (906 IEDs) for TDD ratio $7:8$. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 10 | 1 | WiMAX | 522 / 0 | 422 / 0 | 466.28 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 10 | 1 | Wi-Fi | 243 / 0 | 147 / 0 | 209.59 / 0 | 3 / 0 | 0 / 0 | 1.92 / 0 |
| 15 | 1 | WiMAX | 441 / 0 | 354 / 0 | 410.18 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | Wi-Fi | 105 / 0 | 9 / 0 | 59.15 / 0 | 1 / 0 | 0 / 0 | 0.62 / 0 |
| 20 | 1 | WiMAX | 290 / 0 | 152 / 0 | 229.52 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | Wi-Fi | 18 / 0 | 3 / 0 | 9.05 / 0 | 1 / 0 | 0 / 0 | 0.75 / 0 |
| 30 | 1 | WiMAX | 111 / 0 | 1 / 0 | 54.03 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 1 | Wi-Fi | 15 / 0 | 3 / 0 | 9.12 / 0 | 2 / 0 | 0 / 0 | 1.4 / 0 |
| 30 | 2 | WiMAX | 130 / 0 | 46 / 0 | 65.51 / 0 | 1 / 0 | 0 / 0 | 0.66 / 0 |
| 30 | 2 | Wi-Fi | 19 / 0 | 5 / 0 | 11.73 / 0 | 1 / 0 | 0 / 0 | 0.69 / 0 |
| 40 | 1 | WiMAX | 38 / 0 | 0 / 0 | 8.36 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 10 / 0 | 2 / 0 | 6.46 / 0 | 1 / 0 | 0 / 0 | 0.65 / 0 |
| 40 | 2 | WiMAX | 41 / 0 | 0 / 0 | 15.82 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | Wi-Fi | 11 / 0 | 1 / 0 | 6.61 / 0 | 1 / 0 | 0 / 0 | 0.69 / 0 |

Table 6.10: Number of IEDs with insufficient QoS in Scenario $B$ (906 IEDs) for TDD ratio $10:5$. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 10 | 1 | WiMAX | 430 / 0 | 267 / 0 | 354.83 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 10 | 1 | Wi-Fi | 61 / 0 | 1 / 0 | 9.51 / 0 | 1 / 0 | 0 / 0 | 0.75 / 0 |
| 15 | 1 | WiMAX | 345 / 0 | 201 / 0 | 278.91 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | Wi-Fi | 8 / 0 | 1 / 0 | 3.65 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | WiMAX | 51 / 0 | 0 / 0 | 3.83 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | Wi-Fi | 9 / 0 | 0 / 0 | 3.6 / 0 | 2 / 0 | 0 / 0 | 1.22 / 0 |
| 30 | 1 | WiMAX | 0 / 1 | 0 / 0 | 0 / 0.66 | 0 / 1 | 0 / 0 | 0 / 0.66 |
| 30 | 1 | Wi-Fi | 7 / 0 | 0 / 0 | 2.99 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | WiMAX | 68 / 0 | 0 / 0 | 32.73 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | Wi-Fi | 8 / 0 | 0 / 0 | 3.47 / 0 | 1 / 0 | 0 / 0 | 0.65 / 0 |
| 40 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 5 / 0 | 0 / 0 | 1.66 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | Wi-Fi | 7 / 0 | 0 / 0 | 3.32 / 0 | 1 / 0 | 0 / 0 | 0.62 / 0 |

Table 6.11: Number of IEDs with insufficient QoS in Scenario $B$ (906 IEDs) for TDD ratio 12 : 3. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 10 | 1 | WiMAX | 319 / 0 | 170 / 0 | 208.96 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 10 | 1 | Wi-Fi | 5 / 0 | 0 / 0 | 1.79 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | WiMAX | 225 / 0 | 73 / 0 | 155.78 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | Wi-Fi | 4 / 0 | 0 / 0 | 1.87 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | WiMAX | 25 / 0 | 0 / 0 | 0.97 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | Wi-Fi | 4 / 0 | 0 / 0 | 1.77 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 1 | WiMAX | 1 / 0 | 0 / 0 | 0.19 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 1 | Wi-Fi | 3 / 0 | 0 / 0 | 1.19 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | WiMAX | 2 / 0 | 0 / 0 | 0.12 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | Wi-Fi | 4 / 0 | 0 / 0 | 1.19 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 3 / 0 | 0 / 0 | 1.02 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | Wi-Fi | 2 / 0 | 0 / 0 | 0.76 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |

Table 6.12: Number of IEDs with insufficient QoS in Scenario $B$ (906 IEDs) for TDD ratio $14:1$. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 10 | 1 | WiMAX | 212 / 0 | 170 / 0 | 192.42 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 10 | 1 | Wi-Fi | 2 / 0 | 0 / 0 | 0.56 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | WiMAX | 115 / 0 | 27 / 0 | 81.36 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 15 | 1 | Wi-Fi | 1 / 0 | 0 / 0 | 0.09 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 20 | 1 | Wi-Fi | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 1 | Wi-Fi | 2 / 0 | 0 / 0 | 0.55 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.25 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 30 | 2 | Wi-Fi | 4 / 0 | 0 / 0 | 0.82 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 3 / 0 | 0 / 0 | 0.76 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.01 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 2 | Wi-Fi | 1 / 0 | 0 / 0 | 0.27 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |

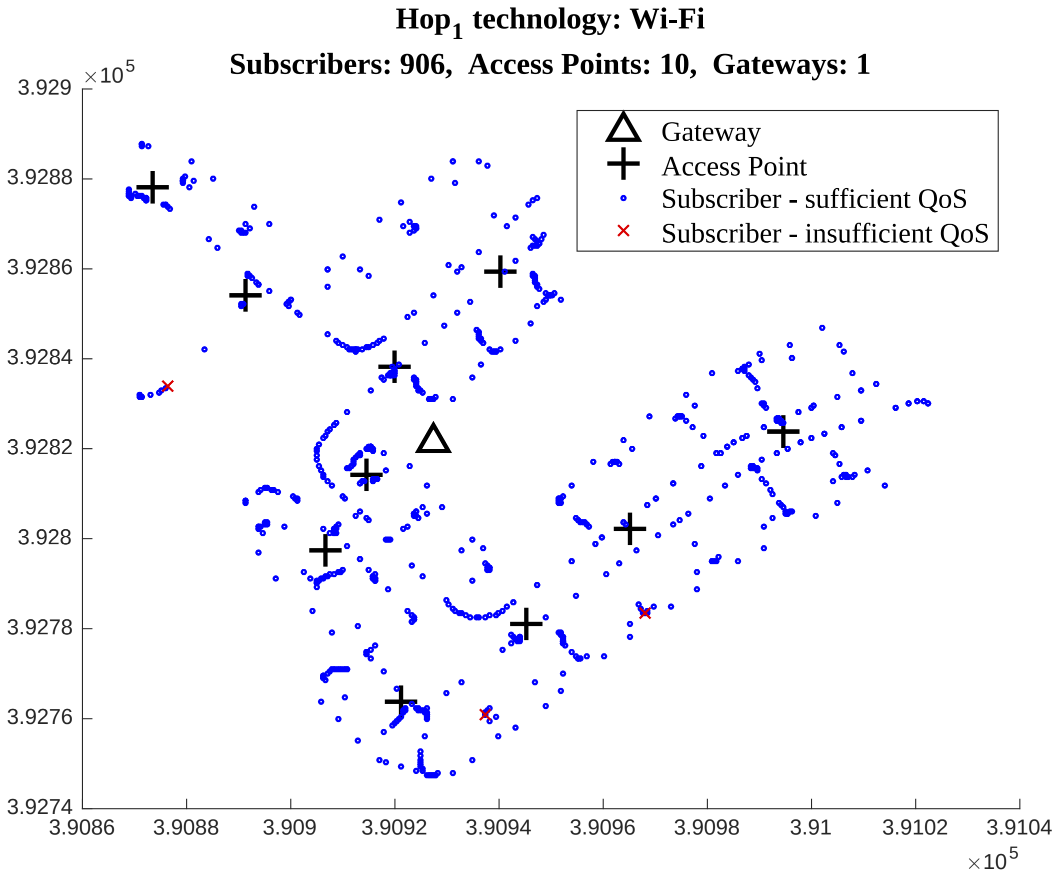



Figure 6.5: Resulting positions of network equipment relative to the 906 subscribers of Scenario $B$. QoS satisfaction results are given for a regular mode of upstream communication with 12 : 3 TDD ratio.

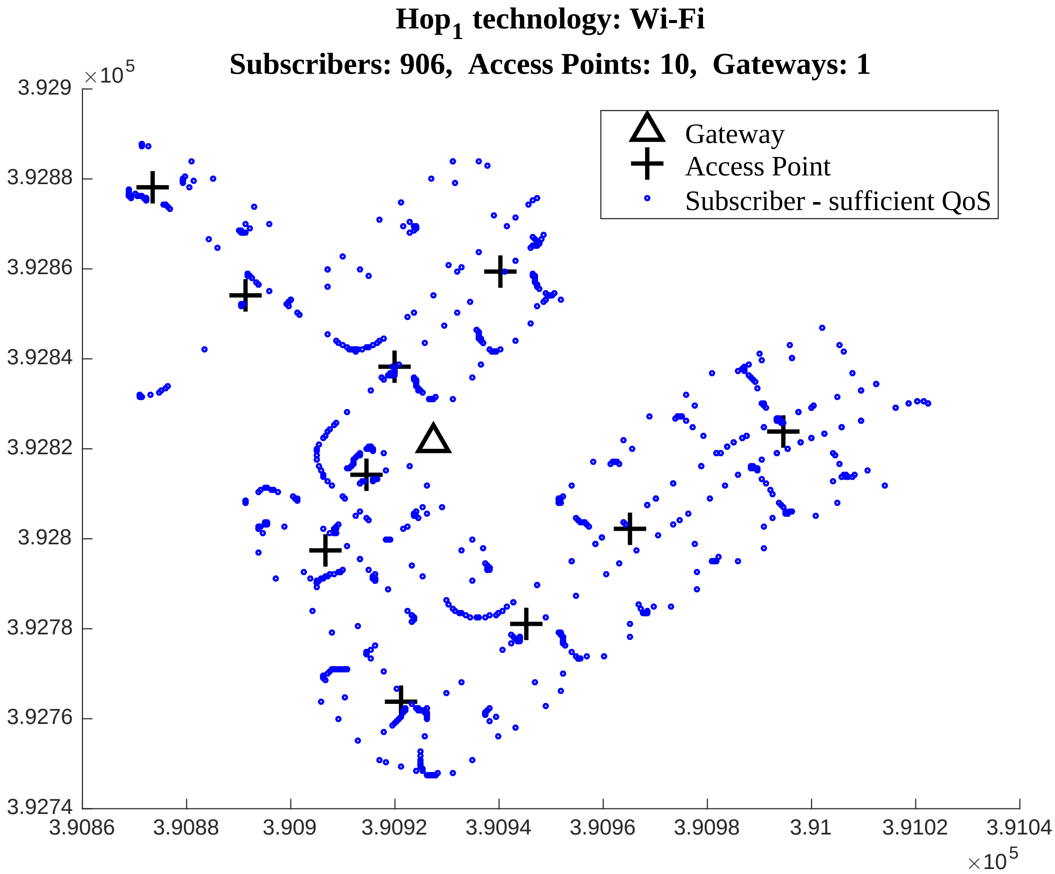



Figure 6.6: Resulting positions of network equipment relative to the 906 subscribers of Scenario $B$. QoS satisfaction results are given for a restricted mode of upstream communication with $12:3$ TDD ratio.

Table 6.13: Number of IEDs with insufficient QoS in Scenario $C$ (2469 IEDs) for TDD ratio $7:8$. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 40 | 1 | WiMAX | 1478 / 0 | 1333 / 0 | 1405.61 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 1420 / 0 | 1243 / 0 | 1325.61 / 0 | 8 / 0 | 1 / 0 | 5.16 / 0 |
| 60 | 1 | WiMAX | 1095 / 0 | 895 / 0 | 991.11 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 60 | 1 | Wi-Fi | 1040 / 0 | 863 / 0 | 959.86 / 0 | 10 / 0 | 2 / 0 | 6.02 / 0 |
| 80 | 1 | WiMAX | 698 / 0 | 478 / 0 | 596.24 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | Wi-Fi | 620 / 0 | 421 / 0 | 537.75 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 1 | WiMAX | 467 / 61 | 293 / 40 | 384.47 / 51.02 | 0 / 61 | 0 / 40 | 0 / 51.02 |
| 100 | 1 | Wi-Fi | 441 / 0 | 301 / 0 | 374.27 / 0 | 8 / 0 | 2 / 0 | 5.13 / 0 |
| 100 | 2 | WiMAX | 219 / 0 | 62 / 0 | 147.02 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 2 | Wi-Fi | 184 / 0 | 72 / 0 | 115.5 / 0 | 3 / 0 | 0 / 0 | 1.95 / 0 |
| 150 | 1 | WiMAX | 254 / 0 | 122 / 0 | 181.67 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 1 | Wi-Fi | 150 / 0 | 90 / 0 | 119.57 / 0 | 10 / 0 | 4 / 0 | 6.66 / 0 |
| 150 | 2 | WiMAX | 99 / 0 | 44 / 0 | 62.89 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 2 | Wi-Fi | 102 / 0 | 68 / 0 | 82.89 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | WiMAX | 55 / 0 | 1 / 0 | 22.51 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | Wi-Fi | 65 / 0 | 24 / 0 | 41.07 / 0 | 4 / 0 | 0 / 0 | 2.61 / 0 |
| 200 | 2 | WiMAX | 23 / 0 | 0 / 0 | 4.79 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | Wi-Fi | 50 / 0 | 25 / 0 | 34.44 / 0 | 2 / 0 | 0 / 0 | 1.3 / 0 |
| 250 | 1 | WiMAX | 9 / 0 | 0 / 0 | 0.46 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 1 | Wi-Fi | 46 / 0 | 21 / 0 | 32.89 / 0 | 2 / 0 | 0 / 0 | 1 / 0 |
| 250 | 2 | WiMAX | 2 / 0 | 0 / 0 | 0.51 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | Wi-Fi | 38 / 0 | 13 / 0 | 24.87 / 0 | 4 / 0 | 0 / 0 | 2.61 / 0 |

Table 6.14: Number of IEDs with insufficient QoS in Scenario $C$ (2469 IEDs) for TDD ratio $10:5$. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 40 | 1 | WiMAX | 1176 / 0 | 979 / 0 | 1085.41 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 1026 / 0 | 863 / 0 | 943.32 / 0 | 3 / 0 | 0 / 0 | 1.85 / 0 |
| 60 | 1 | WiMAX | 514 / 0 | 235 / 0 | 373.44 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 60 | 1 | Wi-Fi | 415 / 0 | 205 / 0 | 292.6 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | WiMAX | 314 / 0 | 166 / 0 | 261.74 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | Wi-Fi | 237 / 0 | 45 / 0 | 122.82 / 0 | 1 / 0 | 0 / 0 | 0.18 / 0 |
| 100 | 1 | WiMAX | 196 / 0 | 82 / 0 | 127.81 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 1 | Wi-Fi | 151 / 0 | 71 / 0 | 117.27 / 0 | 2 / 0 | 0 / 0 | 1.29 / 0 |
| 100 | 2 | WiMAX | 5 / 0 | 0 / 0 | 0.08 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 2 | Wi-Fi | 26 / 0 | 9 / 0 | 17.9 / 0 | 1 / 0 | 0 / 0 | 0.01 / 0 |
| 150 | 1 | WiMAX | 77 / 0 | 39 / 0 | 65.23 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 1 | Wi-Fi | 68 / 0 | 14 / 0 | 39.73 / 0 | 1 / 0 | 0 / 0 | 0.62 / 0 |
| 150 | 2 | WiMAX | 59 / 0 | 41 / 0 | 48.57 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 2 | Wi-Fi | 73 / 0 | 12 / 0 | 52.85 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | WiMAX | 1 / 0 | 0 / 0 | 0.01 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | Wi-Fi | 19 / 0 | 4 / 0 | 11.87 / 0 | 2 / 0 | 0 / 0 | 1.33 / 0 |
| 200 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | Wi-Fi | 19 / 0 | 4 / 0 | 11.4 / 0 | 1 / 0 | 0 / 0 | 0.59 / 0 |
| 250 | 1 | WiMAX | 1 / 0 | 0 / 0 | 0.33 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 1 | Wi-Fi | 14 / 0 | 2 / 0 | 7.36 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.29 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | Wi-Fi | 14 / 0 | 3 / 0 | 8.18 / 0 | 1 / 0 | 0 / 0 | 0.58 / 0 |

Table 6.15: Number of IEDs with insufficient QoS in Scenario $C$ (2469 IEDs) for TDD ratio 12 : 3. In all cases, WiMAX is the communication technology used in $hop_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $hop_1$ | Max | Min | Mean | Max | Min | Mean |
| 40 | 1 | WiMAX | 979 / 0 | 744 / 0 | 869.78 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 831 / 0 | 614 / 0 | 727.61 / 0 | 1 / 1 | / 0 | 0.64 / 0.66 |
| 60 | 1 | WiMAX | 254 / 0 | 122 / 0 | 172.23 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 60 | 1 | Wi-Fi | 226 / 0 | 82 / 0 | 137.08 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | WiMAX | 174 / 0 | 38 / 0 | 99.75 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 80 | 1 | Wi-Fi | 68 / 0 | 8 / 0 | 23.59 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 100 | 1 | WiMAX | 131 / 0 | 15 / 0 | 80.43 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 1 | Wi-Fi | 102 / 0 | 9 / 0 | 39.74 / 0 | 1 / 0 | 0 / 0 | 0.66 / 0 |
| 100 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.01 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 100 | 2 | Wi-Fi | 13 / 0 | 1 / 0 | 7.11 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 150 | 1 | WiMAX | 67 / 0 | 29 / 0 | 40.13 / 0 | 0 / 12 | / 0 | 0 / 1.67 |
| 150 | 1 | Wi-Fi | 31 / 0 | 1 / 0 | 7.5 / 0 | 0 / 0 | / 0 | 0 / 0 |
| 150 | 2 | WiMAX | 60 / 0 | 0 / 0 | 36.87 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 2 | Wi-Fi | 52 / 0 | 0 / 0 | 10.6 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | Wi-Fi | 13 / 0 | 1 / 0 | 6.2 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | Wi-Fi | 9 / 0 | 0 / 0 | 3.39 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 1 | Wi-Fi | 12 / 0 | 1 / 0 | 6.09 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | Wi-Fi | 8 / 0 | 1 / 0 | 3.63 / 0 | 1 / 0 | 0 / 0 | 0.63 / 0 |

Table 6.16: Number of IEDs with insufficient QoS in Scenario $C$ (2469 IEDs) for TDD ratio $14:1$. In all cases, WiMAX is the communication technology used in $\text{hop}_2$.

| | | | Regular (uplink / downlink) | | | Restricted (uplink / downlink) | | |
|---|---|---|---|---|---|---|---|---|
| $k_1$ | $k_2$ | Technology in $\text{hop}_1$ | Max | Min | Mean | Max | Min | Mean |
| 40 | 1 | WiMAX | 842 / 0 | 623 / 0 | 739.54 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 40 | 1 | Wi-Fi | 647 / 0 | 372 / 0 | 527.96 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 60 | 1 | WiMAX | 142 / 0 | 10 / 0 | 81.73 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 60 | 1 | Wi-Fi | 91 / 0 | 7 / 0 | 65.45 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | WiMAX | 79 / 0 | 0 / 0 | 18.22 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 80 | 1 | Wi-Fi | 11 / 0 | 1 / 0 | 4.85 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 1 | WiMAX | 72 / 0 | 0 / 0 | 20.95 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 1 | Wi-Fi | 23 / 0 | 0 / 0 | 3.13 / 0 | 1 / 0 | 0 / 0 | 0.7 / 0 |
| 100 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.07 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 100 | 2 | Wi-Fi | 3 / 0 | 0 / 0 | 0.73 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 1 | WiMAX | 47 / 0 | 0 / 0 | 27.35 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 1 | Wi-Fi | 5 / 0 | 0 / 0 | 1.94 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 2 | WiMAX | 51 / 0 | 0 / 0 | 6.81 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 150 | 2 | Wi-Fi | 2 / 0 | 0 / 0 | 0.52 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 1 | Wi-Fi | 3 / 0 | 0 / 0 | 1.01 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | WiMAX | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 200 | 2 | Wi-Fi | 2 / 0 | 0 / 0 | 0.48 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 1 | WiMAX | 0 / 5 | 5 / 0 | 0 / 2.72 | 0 / 5 | 0 / 0 | 0 / 2.72 |
| 250 | 1 | Wi-Fi | 4 / 0 | 0 / 0 | 1.16 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | WiMAX | 1 / 0 | 0 / 0 | 0.26 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |
| 250 | 2 | Wi-Fi | 2 / 0 | 0 / 0 | 0.55 / 0 | 0 / 0 | 0 / 0 | 0 / 0 |

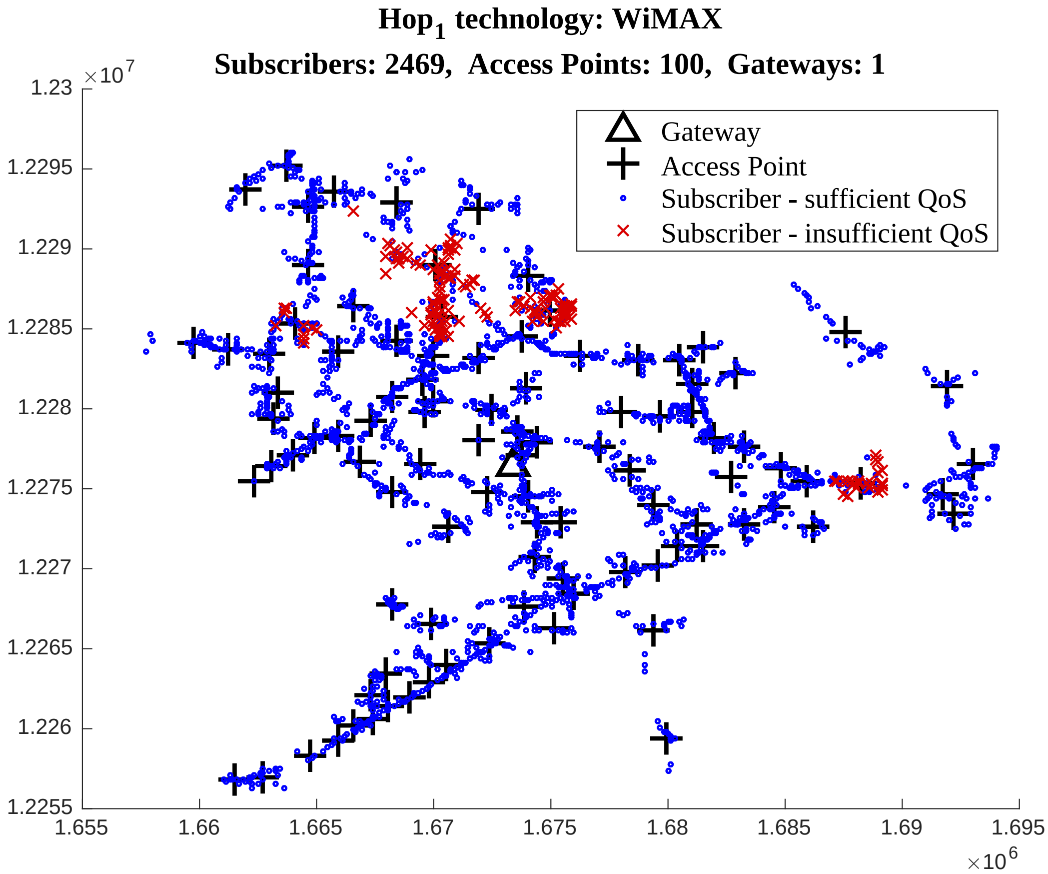



Figure 6.7: Resulting positions of network equipment relative to the 2469 subscribers of Scenario $C$. QoS satisfaction results are given for a regular mode of upstream communication with 10 : 5 TDD ratio.

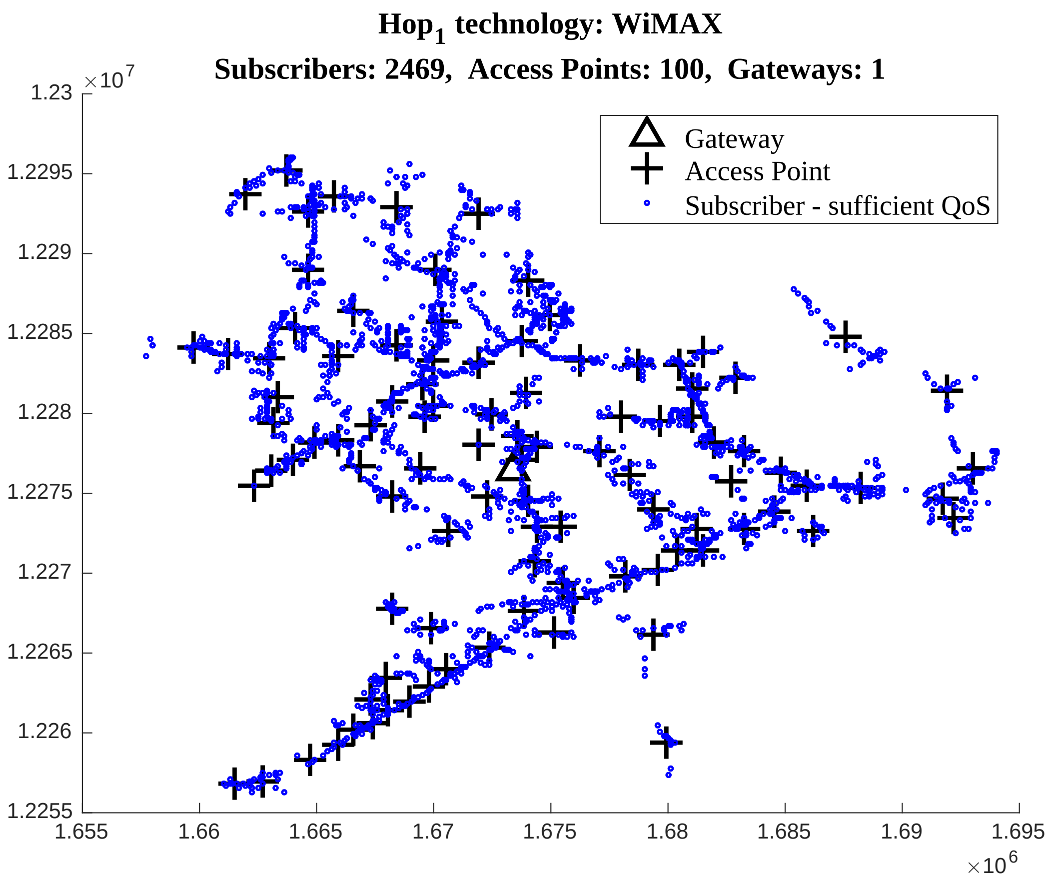



Figure 6.8: Resulting positions of network equipment relative to the 2469 subscribers of Scenario $C$. QoS satisfaction results are given for a restricted mode of upstream communication with $10:5$ TDD ratio.

# Chapter 7

# Conclusions, Contributions, Future Work

## 7.1 Conclusions

As stated in Chapter 1, the purpose of this research is to improve electrical grids and similar networks and inspire innovation in the field. For that reason, smart grid was studied to understand the underlying system and the milestones outlined by the research community. A goal was then set to analyze and optimize data communication in smart grids, with the initial hypothesis stating that the deployment of network cells for wireless communication could be optimized by applying computer science techniques.

Next, several approaches, including graph flow based ones, were studied and the use of clustering techniques was identified as the final go-to approach. It promised good results, as well as interpretability and a variety of algorithms to choose from. Literature in this field was studied and identified a research gap in this direction. This initiated the development of the first version of the proposed model. This initial work used one layer of gateways and compared four popular clustering algorithms for this application.

This was followed by a second model iteration. After literature analysis, development of the second version of the model started, meaning an almost complete redesign of the software project. Imperfections of the project were removed, which led to better approximation of real-world conditions. Lessons learned from the development of the first model helped to guide the development of the final model, resulting in a more accurate, more customizable, and better performing tool.

While the former model compared various clustering algorithms for this domain, the latter studies the use of two types of network cells and uses the most suited clustering algorithm. Both models form a topology graph and calculate data transfer rates for every connection in the topology, which is then used to obtain the communication delays.

A variety of scenario configurations and their results is compared in Chapter 6, reaching a conclusion that wireless technologies can be used for smart grid communications. The used scenarios showed that it is possible to get very high QoS with one gateway per scenario and one AP per about 25 to 100 subscribers.

## 7.2 Contributions

In summary, this work contributes in the following ways to the related fields, particularly with the second iteration of the proposed model:

- In the design phase of the smart grid deployment, a utility company can use the model developed in this thesis to plan the data network, evaluating its feasibility, verifying the impact of possible changes in the data network, using the results as an aid to calculate CAPEX and cost of operation (OPEX) of the data network, and analyze various wireless technologies to use, finally, get insight on where the network access points and gateways should be installed;
- There also exist other non-mobile networks other than that in smart grid that have the same or similar characteristics. The proposed methodology can be applied to any network that meets the following criteria: (i) nodes are non-mobile and their positions are known; (ii) each node has pre-determined static amount of data to transfer in uplink and downlink, coupled with the respective maximum allowed delays;
- The methodology is able to consider that communication happens simultaneously for all nodes. Moreover, both uplink and downlink communication can occur at

the same time. Both of these effects are given as a parameter and influence the communication according to the specified probability. Changes of these characteristics model interference and data traffic and enable studying of the best-case and worst-case scenario regarding scheduling, or anything in between;

- The resulting methodology is modifiable to a fairly high degree, making it a ubiquitous tool for QoS modeling in general. This can be demonstrated by the benefits described above and by the amount of parameters used in Chapter 6;

- A mode of communication restriction was studied in the thesis to show how to model the effect of temporary communication changes in the network on QoS, which is useful when the network capacity is insufficient, relative to its traffic;

- The source code of the two-level methodology implementation is publicly available under an open-source license online[1]. Using this tool, researchers and companies can model the delay QoS behavior based on their scenario description, using the proposed methodology. Most importantly, the model is also configurable. Coordinates of new and completely different scenarios can be plugged into the tool to create a model to represent the scenario and study the network necessary to deploy;

- One article was published as a product of this work. Specifically, the comparison of clustering algorithms with the use of the single-level methodology:

  - Vrbský, L., da Silva, M. S., Cardoso, D. L., & Francês, C. R. L. (2017, May). Clustering techniques for data network planning in Smart Grids. In *Networking, Sensing and Control (ICNSC), 2017 IEEE 14th International Conference on* (pp. 7-12). IEEE.

## 7.3 Future Work

There are several ways to continue the work on this model or in this direction in general. For instance, the locations of installed network equipment (access points and gateways) can be further optimized to find better local optima than what the current

[1] `https://github.com/vrbsky/network-model`

model finds. One opportunity is to take into account the inverse-square law of propagation, which says that the signal strength between two nodes is inversely proportional to the square of the distance between them [Sandnes et al., 2008]. One example of such modification is to use the cluster separation found by the clustering algorithm and then re-compute the optimal position of the cluster head of each cluster based on the quadratic relationship.

The model can also be adapted to a mobile scenario, where the input coordinates to determine the positions of the network equipment would represent the probability distribution of active user devices in the scenario. A different set of coordinates, representing the actual positions of users at a given time, would then be used to calculate the actual transfer delays, while keeping the cell positions fixed. These two sets of coordinates can even have different number of elements if a minor alteration of the source code is performed.

Another opportunity is to use other distance measures that better express distance in a network. For wireless networks that could be a square of the Euclidean distance for example. Such measures can be expressed in form of a kernel function during clustering.

Also, there is a possibility to study the reconfiguration of the network itself, as well as just its topology, in case of a partial network failure. Such concept would be applicable in the occurrence of a partial energy blackout or natural disasters. In such case the network would be capable of reconfiguring the topology to satisfy the QoS of the whole network. In some cases, deployment of temporary network nodes would apply, before a final solution is found and carried out.

Finally, it is possible to let an algorithm manipulate the configurations of the model. This way, another layer of optimization would be deployed. One possibility is to use algorithms like genetic algorithms or a neural networks to manipulate the cell counts between the model executions.

# Bibliography

3GPP (2012). Universal mobile telecommunications system (umts); physical channels and mapping of transport channels onto physical channels (tdd) (3gpp ts 25.221 version 11.0.0 release 11). Technical report, 3GPP.

Aggarwal, C. and Reddy, C. (2016). *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press.

Alcatel-Lucent (2010). Technology white paper - smart choices for the smart grid. Technical report, Alcatel-Lucent Enterprise.

Arnold, G. W. (2011). Challenges and opportunities in smart grid: A position article. *Proceedings of the IEEE*, 99(6):922–927.

Arritt, R. and Dugan, R. (2010). The ieee 8500-node test feeder. In *IEEE PES T&D 2010*, pages 1–6. IEEE.

Asbery, C. W., Jiao, X., and Liao, Y. (2016). Implementation guidance of smart grid communication. In *North American Power Symposium (NAPS), 2016*, pages 1–6. IEEE.

Bakken, D. E., Bose, A., Hauser, C. H., Whitehead, D. E., and Zweigle, G. C. (2011). Smart generation and transmission with coherent, real-time data. *Proceedings of the IEEE*, 99(6):928–951.

Brunner, C. (2008). Iec 61850 for power system communication. In *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, pages 1–6.

Budka, K. C., Deshpande, J. G., Thottan, M., et al. (2014). *Communication networks for smart grids: making smart grid real*. Springer.

Chakraborty, M. and Chaki, N. (2015). An ipv6 based hierarchical address configuration scheme for smart grid. In *Applications and Innovations in Mobile Computing (AIMoC), 2015*, pages 109–116. IEEE.

Chatzimisios, P., Stratogiannis, D., Tsiropoulos, G., and Stavrou, G. (2013). A survey on smart grid communications: from an architecture overview to standardization activities. *Handbook of Green Information and Communication Systems.*

Cheng, S. and Day, M. (2014). *Technologies and Applications of Artificial Intelligence: 19th International Conference, TAAI 2014, Taipei, Taiwan, November 21-23, 2014, Proceedings*. Lecture Notes in Computer Science. Springer International Publishing.

CIGRE WG34.03 (2001). Communication requirements in terms of data flow within substations. Technical Report Technical Brochure 180, CIGRE WG34.03.

Curtis, K. (2005). A dnp3 protocol primer. *DNP User Group*, 2005.

Dhillon, H. S., Ganti, R. K., and Andrews, J. G. (2011). A tractable framework for coverage and outage in heterogeneous cellular networks. In *Information Theory and Applications Workshop (ITA), 2011*, pages 1–6. IEEE.

DOE (2012). 2010 smart grid system report. Technical report, U.S. Department of Energy.

DOE (2013). Economic impact of recovery act investments in the smart grid.

DOE (2018). Table 1.1. net generation by energy source: Total (all sectors), 2006-september 2016. Technical report, U.S. Department of Energy.

EISA (2007). Security act (eisa). In *110th United States Congress. Energy independence and security act of*, volume 2007.

Ekici, E. and Ersoy, C. (2001). Multi-tier cellular network dimensioning. *Wireless Networks*, 7(4):401–411.

Emmanuel, M. and Rayudu, R. (2016). Communication technologies for smart grid applications: A survey. *Journal of Network and Computer Applications*, 74:133–148.

Erol-Kantarci, M. and Mouftah, H. T. (2015). Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues. *IEEE Communications Surveys & Tutorials*, 17(1):179–197.

ETP (2010). Smartgrids strategic deployment document for europe's electricity networks of the future. *European Technology Platform SmartGrids*.

Everitt, B. S., Landau, S., and Leese, M. (2001). Cluster analysis arnold. *A member of the Hodder Headline Group, London*.

Fan, Z., Kulkarni, P., Gormus, S., Efthymiou, C., Kalogridis, G., Sooriyabandara, M., Zhu, Z., Lambotharan, S., and Chin, W. H. (2013). Smart grid communications: Overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys & Tutorials*, 15(1):21–38.

Fang, X., Misra, S., Xue, G., and Yang, D. (2012). Smart grid-the new and improved power grid: A survey. *IEEE communications surveys & tutorials*, 14(4):944–980.

Farhangi, H. (2010). The path of the smart grid. *IEEE power and energy magazine*, 8(1).

Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *science*, 315(5814):972–976.

Gellings, C. (2011). Estimating the costs and benefits of the smart grid: a preliminary estimate of the investment requirements and the resultant benefits of a fully functioning smart grid. *Electric Power Research Institute (EPRI), Technical Report (1022519)*.

Gong, Y. and Guzmán, A. (2011). Distribution feeder fault location using ied and fci information. In *Protective Relay Engineers, 2011 64th Annual Conference for*, pages 168–177. IEEE.

Gungor, V., Sahin, D., Kocak, T., and Ergüt, S. (2011a). Smart grid communications and networking. *Turk Telecom*.

Gungor, V. C. and Lambert, F. C. (2006). A survey on communication networks for electric system automation. *Computer Networks*, 50(7):877–897.

Gungor, V. C., Lu, B., and Hancke, G. P. (2010). Opportunities and challenges of wireless sensor networks in smart grid. *IEEE transactions on industrial electronics*, 57(10):3557–3564.

Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., and Hancke, G. P. (2011b). Smart grid technologies: Communication technologies and standards. *IEEE transactions on Industrial informatics*, 7(4):529–539.

Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., and Hancke, G. P. (2013). A survey on smart grid potential applications and communication requirements. *IEEE Transactions on Industrial Informatics*, 9(1):28–42.

Guo, L., Dong, M., Ota, K., Wu, J., and Li, J. (2016). A name-based secure communication mechanism for smart grid employing wireless networks. In *Global Communications Conference (GLOBECOM), 2016 IEEE*, pages 1–6. IEEE.

Hauser C. H., Bakken D. E., B. A. (2005). A failure to communicate: next generation communication requirements, technologies, and architecture for the electric power grid. *IEEE Power and Energy Magazine*, 3(2):45–55.

Heo, N. and Varshney, P. K. (2003). An intelligent deployment and clustering algorithm for a distributed mobile sensor network. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 5, pages 4576–4581. IEEE.

IEC 61850 (2003). Iec 61850: Communication networks and systems in substations. *Int. Electrotech. Commission, Geneva, Switzerland.*

IEEE 1815 (2012). Ieee standard for electric power systems communications-distributed network protocol (dnp3). *IEEE Std 1815-2012 (Revision of IEEE Std 1815-2010)*, pages 1–821.

IEEE PES (2015). The ieee european low voltage test feeder. Technical report, IEEE Power & Energy Society.

Islam, M. and Lee, H.-H. (2012). Iec61850 based operation, control and management of utility connected microgrid using wireless technology. In *International Conference on Intelligent Computing*, pages 258–265. Springer.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Júlio, G. F. A. (2015). Algoritmo para planejamento e otimização de topologia de redes de comunicação de dados para smart grid. Bachelor thesis, Universidade Federal do Pará.

Kabalci, Y. (2016). A survey on smart metering and smart grid communication. *Renewable and Sustainable Energy Reviews*, 57:302–318.

Kawadia, V. and Kumar, P. (2003). Power control and clustering in ad hoc networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 1, pages 459–469. IEEE.

Kersting, W. (2012). *Distribution System Modeling and Analysis, Third Edition*. Taylor & Francis.

Kersting, W. H. (2001). Radial distribution test feeders. In *Power Engineering Society Winter Meeting, 2001. IEEE*, volume 2, pages 908–912. IEEE.

Kheaksong, A., Prayote, A., and Lee, W. (2016). Performance evaluation of smart grid communications via network simulation version 3. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on*, pages 1–5. IEEE.

Kitchen, R. (2001). *RF and Microwave Radiation Safety Handbook*. Electronics & Electrical. Newnes.

Laverty, D. M., Morrow, D. J., Best, R., and Crossley, P. A. (2010). Telecommunications for smart grid: Backhaul solutions for the distribution network. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–6. IEEE.

Lin, C. R. and Gerla, M. (1997). Adaptive clustering for mobile wireless networks. *IEEE Journal on Selected areas in Communications*, 15(7):1265–1275.

Liu, Y., Gao, H., Wei, X., Teng, Z., Wei, P., Li, N., and Xiang, M. (2011). Study on new type of ied with integrated functions in intelligent substation. In *Advanced*

*Power System Automation and Protection (APAP), 2011 International Conference on*, volume 1, pages 161–165. IEEE.

Mahmood, A., Javaid, N., and Razzaq, S. (2015). A review of wireless communications for smart grid. *Renewable and sustainable energy reviews*, 41:248–260.

McGranaghan, M., Dollen, D. V., Myrda, P., and Gunther, E. (2008). Utility experience with developing a smart grid roadmap. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–5.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill international editions - computer science series. McGraw-Hill Education.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to linear regression analysis*, volume 821. John Wiley & Sons.

Müller, C., Georg, H., and Wietfeld, C. (2012). A modularized and distributed simulation environment for scalability analysis of smart grid ict infrastructures. In *Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques*, pages 327–330. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Nasreddine, J. and Lagrange, X. (2003). Time slot allocation based on a path gain division scheme for td-cdma tdd systems. In *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, volume 2, pages 1410–1414. IEEE.

NETL (2008). Advanced metering infrastructure. *National Energy Technology Laboratory for the US Department of Energy Office of Electricity and Energy Reliability.*

NIST 1.0 (2010). Nist framework and roadmap for smart grid interoperability standards, release 1.0. *National Institute of Standards and Technology*, 33.

NIST 3.0 (2014). Nist framework and roadmap for smart grid interoperability standards, release 3.0. *National Institute of Standards and Technology (NIST), Special Publication 1108R3*, pages 2–0.

Niyato, D., Dong, Q., Wang, P., and Hossain, E. (2013). Optimizations of power consumption and supply in the smart grid: Analysis of the impact of data communication reliability. *IEEE Transactions on Smart Grid*, 4(1):21–35.

Niyato, D., Wang, P., Han, Z., and Hossain, E. (2011). Impact of packet loss on power demand estimation and power supply cost in smart grid. In *2011 IEEE Wireless Communications and Networking Conference*, pages 2024–2029. IEEE.

Noorwali, A., Rao, R., and Shami, A. (2016). Modeling and delay analysis of wide area network in smart grid communications. In *Smart Energy Grid Engineering (SEGE), 2016 IEEE*, pages 347–352. IEEE.

Pavan, M. and Pelillo, M. (2003). A new graph-theoretic approach to clustering and segmentation. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–145. IEEE.

Qingling, Z. and Li, J. (2006). Rain attenuation in millimeter wave ranges. In *Antennas, Propagation & EM Theory, 2006. ISAPE'06. 7th International Symposium on*, pages 1–4. IEEE.

Rahnamay-Naeini, M., Wang, Z., Mammoli, A., and Hayat, M. M. (2012). Impacts of control and communication system vulnerabilities on power systems under contingencies. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–7. IEEE.

Rappaport, T. S., Murdock, J. N., and Gutierrez, F. (2011). State of the art in 60-ghz integrated circuits and systems for wireless communications. *Proceedings of the IEEE*, 99(8):1390–1436.

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., and Edwards, D. D. (2003). *Artificial intelligence: a modern approach*, volume 2. Prentice hall Upper Saddle River.

Sandnes, F., Zhang, Y., Rong, C., and Yang, L. (2008). *Ubiquitous Intelligence and Computing: 5th International Conference, UIC 2008, Oslo, Norway, June 23-25, 2008 Proceedings*. Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Silva, M. S. d. (2014). *Estratégia de planejamento e otimização de redes de comunicação de dados como suporte à implantação de smart grids*. PhD thesis, Universidade Federal do Pará.

Sooriyabandara, M. and Ekanayake, J. (2010). Smart grid-technologies for its realisation. In *2010 IEEE International Conference on Sustainable Energy Technologies (ICSET)*, pages 1–4. IEEE.

Taranetz, M. and Rupp, M. (2012). Performance of femtocell access point deployments in user hot-spot scenarios. In *Telecommunication Networks and Applications Conference (ATNAC), 2012 Australasian*, pages 1–5. IEEE.

Vallejo, A., Zaballos, A., Selga, J. M., and Dalmau, J. (2012). Next-generation qos control architectures for distribution smart grid communication networks. *IEEE Communications Magazine*, 50(5):128–134.

World Bank (2017). *World Development Indicators 2017*. World Bank.

WWF (2010). The energy report. Technical report, World Wide Fund for Nature.

Xing, N., Zhang, S., Shi, Y., and Guo, S. (2016). Plc-oriented access point location planning algorithm in smart-grid communication networks. *China Communications*, 13(9):91–102.

Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678.

Yan, Y., Qian, Y., Sharif, H., and Tipper, D. (2013). A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE communications surveys & tutorials*, 15(1):5–20.

# Appendix A

# SINR Calculation

The calculation of SINR and all of its parts, considered in this thesis, is presented in this section.

First, let $\Gamma \in \{\text{uplink}, \text{downlink}\}$ denote the communication direction considered in the current SINR calculation, i.e. in the direction from $i$ to $j$. $P_{i,j}$ denotes the power received by $j$ from the equipment $i$, defined as

$$P_{i,j}(\text{dBm}) = P_i + G_i + G_j - L_{i,j}\,, \tag{A.1}$$

where $L_{i,j}$ is the path loss between $i$ and $j$ in dB, $P_i$ is the transmitting power of $i$ in dBm, $G_i$ and $G_j$ is the gain of the antenna of $i$ and $j$ in dB, respectively. There are various radio wave propagation models that define specific path loss and its calculation. Free Space Path Loss that is used in this thesis, is defined as follows:

$$\begin{aligned} L_{i,j}(\text{dB}) = 92.4 + 20\log_{10}(d_{i,j}) + 20\log_{10}(f_{i,j}) + \\ + \alpha_{\text{atm}}(d_{i,j}) + \alpha_{\text{rain}}(d_{i,j}) + \varphi_{i,j} + h_{i,j}\,, \end{aligned} \tag{A.2}$$

where $f$ is the carrier frequency in GHz, and $d_{i,j}$ stands for the Euclidean distance between $i$ and $j$ in kilometers, given as

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\,, \tag{A.3}$$

where $(x_i, y_i)$ represents the Cartesian coordinates of some node $i$. $\alpha_{\text{atm}}$ and $\alpha_{\text{rain}}$ is the atmospheric and rain attenuation respectively, and $\varphi_{i,j}$ stands for foliage losses:

$$\varphi(i,j) = 0.2\,(f_{i,j})^{0.3}(R)^{0.6}\,, \tag{A.4}$$

where $R$ is the depth of the foliage in meters and $f_{i,j}$ is the operating frequency used between $i$ and $j$, in MHz. Also, let $f_\Gamma, a$ be the operating frequency of network node $a$ in the direction $\Gamma$. Next, let $b_i \in \{1, 2, \ldots, |\text{bands}|\}$ be the band used by $i$, and $\text{rand}(x, y)$ be a function that draws a random number from uniform distribution over the interval $[x, y]$. The probability of the equipment of any node being active is denoted by $\mathrm{p}_\Gamma$. Then, for a given TDD ratio u : d,

$$\mathrm{p}_\Gamma = \begin{cases} \frac{\mathrm{u}}{\mathrm{u+d}}\,, & \text{if } \Gamma = \text{uplink;} \\ \frac{\mathrm{d}}{\mathrm{u+d}}\,, & \text{otherwise.} \end{cases} \tag{A.5}$$

With that, let $K_{i,j}$ be a set of all interfering nodes with communication from $i$ to $j$, given as

$$a \in K_{i,j} \Leftrightarrow \big(f_{\Gamma,a} = f_{j,i} \wedge \left|b_a - b_j\right| \leq 1 \wedge \text{rand}(0,1) \leq \mathrm{p}_\Gamma\big), \forall a \in A \setminus \{i,j\}\,, \tag{A.6}$$

where $A$ is the set of all network nodes.

Let $I_{i,j}$ denote the total interference at $j$, when receiving data from $i$. Assuming perfect orthogonality of an OFDMA-based network or other spectrum management method, the interference at any $j$ in this thesis depends only on the inter-cell interference. With that, the interference can be calculated as

$$I_{i,j} = \sum_{k \in K_{i,j}} P_{k,j}\,. \tag{A.7}$$

In this thesis, noise is denoted as $N$ and is considered to be equal to thermal noise $N_T$, given as

$$N_T(\mathrm{W}) = 10\log_{10}(k_{\mathrm{B}} \times T \times 1000) + 10\log_{10}(\beta)\,, \tag{A.8}$$

where $k_B$ is Boltzmann constant with the value of $1.380\,648\,52 \times 10^{-16}\,\mathrm{J\,K^{-1}}$, $T$ is the ambient temperature in Kelvin, and $\beta$ is the communication bandwidth in Hz.

Finally, with the above defined, let $\mathrm{SINR}_{i,j}$ be the SINR of the channel from a node $i$ to a node $j$, given as

$$\mathrm{SINR}_{i,j} = \frac{P_{i,j}}{I_{i,j} + N}\,, \tag{A.9}$$

where all three quantities on the right-hand side of the equation are expressed in Watts.