FEDERAL UNIVERSITY OF PARÁ
INSTITUTE OF TECHNOLOGY
POST-GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

NAGIB COELHO MATNI NETO

FUZZY C-MEANS BASED GATEWAY PLACEMENT ALGORITHM FOR LORAWAN

DM: 12/2020

UFPA / ITEC / PPGEE
Guamá University Campus
Belém-PA

February / 2020

NAGIB COELHO MATNI NETO

FUZZY C-MEANS BASED GATEWAY PLACEMENT ALGORITHM FOR LORAWAN

Master Thesis submitted to the judging panel at the Federal University of Pará as part of the requirements for obtaining a Master's degree in Eletrical Engineering

Advisor: Denis Lima do Rosário

Co-Advisors: Eduardo Coelho Cerqueira

BELÉM-PA

February / 2020

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

M425f Matni, Nagib
FUZZY C-MEANS BASED GATEWAY PLACEMENT ALGORITHM FOR LORAWAN / Nagib Matni. — 2020.
xv,50 f. : il. color.

Orientador(a): Prof. Dr. Denis Lima RosÁrio
Coorientador(a): Prof. Dr. Eduardo Coelho Cerqueira
Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Tecnologia, Universidade Federal do Pará, Belém, 2020.

1. LPWAN. 2. IoT. 3. Gateway Placement. 4. CAPEX. 5. OPEX. I. Título.

CDD 621.3

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

"FUZZY C-MEANS BASED GATEWAY PLACEMENT ALGORITHM FOR LORAWAN"

AUTOR: NAGIB COELHO MATNI NETO

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: 27/02/2020

BANCA EXAMINADORA:

Prof. Dr. Denis Lima do Rosário
(Orientador – PPGEE/UFPA)

Prof. Dr. Eduardo Coelho Cerqueira
(Co-Orientador – PPGEE/UFPA)

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

Prof. Dr. Helder May Nunes de Oliveira
(Avaliador Externo ao Programa – ICEN/UFPA)

VISTO:

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

Dedicated to my grandmother Lycia, my parents, my girlfriend, and friends who gave me strength and word of encouragement.

Acknowledgments

First of all, I thank God, for my existence, for the protection and comfort in times of difficulty. To my parents and my sister, for all the love, care and education, being inspiration in my day to day, and foundation that allowed security in the search for knowledge that would contribute in my formation. I would also like to thank my girlfriend, Luciana, for have a huge patience and for all "I can't go, I have to study", for helping with corrections, and especially for the companionship.

To my Advisor and Co-Advisor, Prof. Dr. Denis Rosário and Eduardo Cerqueira, always present with orientations and dialogues, contributing to my professional and personal formation. A special thanks goes to GERCOM (UFPA), who supported, helped and gave useful suggestions to assemble the parts in the development process,especially to Jean, who helped me constantly with the simulations. Last but not least, I appreciate the guidance given by other supervisors in the project presentation that has improved my presentation skills, thanks to their comments and advice.

In addition, this work was conducted during a scholarship supported by the International Cooperation Program CAPES at the Federal University of Pará. Financed by CAPES–Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil, and by the R&D project entitled "IoT-Cloud de Medição de Energia Centralizada Voltado a Rede CEA", process number 001/2017.

Abstract

Abstract of the master thesis presented to the jury as a partial fulfillment of the requirements for the Master's degree in the Postgraduate Program of Eletrical Engineering.

Fuzzy C-Means Based Gateway Placement Algorithm for LoRaWAN

Advisor: Denis Lima do Rosário
Co-advisor: Eduardo Coelho Cerqueira
Key words: LPWAN; IoT; LoRaWAN; CAPEX; OPEX; Gateway Placement.

Low Power Wide Area Network (LPWAN) technologies recently gained interest from the research and industrial community. Internet of Things (IoT) devices communicate directly with gateways, which act as bridges towards a central network server and the Internet. In this context, it is important to study how to place multiple gateways in an area considering Quality of Service, Capital expenditure (CAPEX), and operational expenditure (OPEX) requirements. This is because network planning and optimization are considered to be significant issues that impact on the application performance, CAPEX, and OPEX. In this master thesis, we propose an optimal LoRa gateway placement (PLACE). It considers the Gap statistics method to find the number of LoRa gateway, which is used to compute the gateway placement using the Fuzzy C-Means algorithm. Simulation results show that PLACE reduced in 36% the CAPEX and OPEX compared to the grid and random gateway placement, while keeps a similar Packet Delivery Ratio.

List of Figures

List of Tables

List of Acronyms

ADR	Adaptive Data Rate
AWGN	Additive White Gaussian Noise
CAPEX	Capital Expenditure
CR	Coding Tate
CSS	Chirp Spread Spectrum
FCM	fuzzy c-means
FSK	Frequency Shifting Keying
ILP	Integer Linear Programming
IoT	Internet of things
LoRa	Long-Range
LoRaWAN	LoRa Wide-Area Network
LPWA	Low Power Wide Area
LPWAN	Low Power Wide Area Network
MAC	Medium Access Control
MIC	Message Integrity Checks
mMTC	Massive Machine-Type Communication
NF	Noise Figure
OPEX	Operating Expenses
QoS	Quality of Service
ROC	Reliability of Connections
RSSI	Received Signal Strength Indicator
SFD	Start Frame Delimiter
SF	Spreading Factor
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
ToA	Time on Air

List of Symbols

f_{min}	Minimum Frequency
f_{max}	Maximum Frequency
R_s	Symbol Rate
R_c	Chip Rate
T_s	Duration of a Symbol
R_b	Bitrate
i	The Spread Correcting Factor
j	The Transmission Employing Spreading Factor
T_{RX}_DELAY1	The First Reception Window
T_{RX}_DELAY2	The Second Reception Window
T_Beacon	Time Synchronized Beacon
l_u	the path loss in urban areas
h_{rx}	the height of gateway antenna
h_{tx}	the height of device antenna
f	the frequency of transmission
c_h	the antenna height correction factor
d	the distance between the gateway and the device
c	Cluster index
i	Object index
μ_{ic}	The Membership Coefficient
m	Weighting parameter
D_{ic}^2	The Standard Euclidean Distance
$Jm_{c,b}^*$	Fuzzy C-Means Objective Function
μ_{ic}	Membership Percentage
U	C-Partition Matrix with all the Objects Membership
C	The maximum number of clusters

Jm	Fuzzy C-Means Objective Function
τ	The Reporting Periodicity
G_t	Antenna Gain
C_{Bs}	The Cost of a LoRaWan Gateway Acquisition
C_{ins}	The Cost of a Gateway Deployment
C_{set}	The Cost of a Gateway Setup
T_{xinst}	The cost of Transmission Installation
C_{man}	The Operation and Maintenance Cost
C_{lease}	The Lease Cost
C_{elet}	The Electricity Cost per Year
T_{xinst}	The Transmission Cost
t	The time in Years

Contents

CHAPTER 1

Introduction

1.1 Overview

The recent explosion of the Internet of things (IoT) technology is dramatically changing society through disruptive technologies in new sets of verticals and applications [44]. Smart cities, healthcare, agriculture, environmental monitoring, logistics, home/building automation, smart grid, and critical infrastructure monitoring are only footprint examples with regards to smart environment verticals [2, 1, 47] targeting many consumers. Available forecasts estimate that 20.8 billion connected things will exist by 2020 [62], the number of 5G-connected IoT devices will reach 4.1 billion by 2024 [12], and 500 billion IoT devices are expected to be connected to the Internet by 2030 [21]. In light of these high-dense IoT network predictions, new challenging requirements arise, which drive to a global connectivity reassessment [10, 16] to enable enhanced potentials in harnessing real-time planning models.

The list of IoT devices communication requirements is heterogeneous, including low energy consumption (to address 10-years battery life), high coverage, and massive Machine-Type Communication (mMTC) [7]. Indeed, the Internet core is evolving smoothly, driven by high-expanding large-scale broadband capabilities that optical networking technologies offer. On the other hand, access network infrastructures mostly harness short-range, long-range, and cellular network deployments. Although mature, existing network access technologies need to evolve to suit IoT requirements [3].

Capital expenditures (CAPEX), represents investments or spending's in capital goods, which are those used in the production of other items, such as equipment, construction materials, among others. In other words, it is the funds used to acquire elements that will help expand the company's ability to generate profit. Also, Operating expenses (OPEX), which are payments related to business management activities and the sale of products and services. This is the case, for example, of a company that would buy a computer, but change to a service that delivers the machine and still monitors trains employees, and updates the system.

Firstly, widely adopted short-range network technologies (*e.g.,* WiFi and Bluetooth) provide a few meters limited coverage area, and are highly interference-affected. Such networks need highly-dense deployment to achieve expanded coverage area [15, 42]; thus, impacting increased expenditures (CAPEX/OPEX), increasing exposition to noise and interference, as well as very complex management. Traditional cellular networks operating on licensed frequency bands (*i.e.,* 2G, 3G, and 4G) are capable of provisioning extended coverage from tens up to hundreds of meters, attending thousands of devices with high throughput [13]. However, modulation complexity, along with medium access schemes, are highly energy demanding aspects. To capture this growth, innovative long-range technologies, such as Low Power Wide Area Network (LPWAN) [54], is attracting attention from both academy and industry through promising capabilities in broad area connectivity operating on unlicensed frequency bands with the appropriate data rate, power consumption, and throughput tailored to IoT application verticals [48].

In this context, the Long-Range (LoRa) technology is considered the most adopted LPWAN technology [41, 63]. Based on the LoRa physical layer, the LoRa Alliance [4] defines the LoRa Wide-Area Network (LoRaWAN) Medium Access Control (MAC) layer protocol standard [17]. In particular, the LoRa physical layer adopts the robust Chirp Spread Spectrum (CSS) modulation with different Spreading Factor (SF) [37]. On top of the LoRa physical layer, LoRa Alliance defines the higher layers and network architecture [18]. The MAC layer of LoRaWAN is essentially an ALOHA variant of random access owing to its simplicity [20]. Regarding the architecture, there are potentially a high number of IoT devices sending data to the application server via the same LoRa Gateway.

LoRa usually considers one gateway to cover several devices, which could add one more gateway to split the number of devices in order to improve the Quality of Service (QoS) [65]. The densification of LoRaWAN poses coexistence challenges as the deployment of gateways populate urban areas, which brings new challenges to the connectivity protocols that are currently available. However, increasing the number of gateways also increases the costs, and thus balancing the interest in terms of QoS, CAPEX, and OPEX is a challenging task.

1.2 Motivation and challenges

Gateway (also named as “base station” in mobile communication systems) planning is critical to LoRa networks. Currently, many previous studies [14, 26] have optimized the deployment of base stations in mobile communication with different objectives. However, planning LoRa gateways is somewhat different from that in conventional mobile networks. It is challenging from the following two perspectives. Firstly, the LoRa networks emphasize on the reliability of connections (ROC). Unlike mobile communications, IoT devices in LoRa networks such as smoke alarms, as well as access controllers, are generally static. They can not move around to have a proper channel as in mobile communications when the coverage hole exists. In such a case, these IoT devices will be disconnected from the Internet. How to provide reliable connections is a challenge for LoRa network planning.

Secondly, when it involves large-scale inputs (such as thousands of candidate locations), Branch and Bound (B&B) method, which is commonly used to solve small-scale integer programming problems, fails to terminate within acceptable runtime (e.g., several days). Typical approaches such as greedy algorithms do not perform well due to the unique characteristic of our problem: in the two-type gateway planning problem, each candidate location has two options of gateway deployment, but only one gateway can be selected. Such a combinational decision (which one to choose and which type to assign) will have an impact on subsequent decisions, making deployment interdependent.

In this sense, the problem we are studying in this master thesis is how to place multiple LoRa gateways in an area, which has possibly hundreds or thousands of deployed devices. This is because network planning and optimization is an important issue that impacts the QoS, CAPEX, and OPEX [59]. The communication channel between IoT devices and gateway can be significantly improved, as a result of such efficient LoRa gateway placement, while reduces the CAPEX and OPEX. While previous studies have examined placement problems, to the best of our knowledge, none of them has addressed the question of the LoRa gateway placement, including minimum capacity requirements considering cost-efficient. It is common to associate the dilemma of placement gateways with clustering, where the number of gateways refers to the number of clusters. In other words, we try to divide the end devices into groups with some similarity between them, in order to be able to place a gateway that can serve these devices in the best possible way.

The capability of an IoT network protocol to fulfill those above requirements needs to be carefully investigated before a massive deployment can be implemented. Currently, a debate is going on about the adequate performance of many different network standards. One particular architecture for IoT networks, the LPWAN paradigm, is still in the evaluation phase in the research community: the objective is to understand whether these networks are a viable solution for the deployment of massive IoT and whether they will be able to compete with other standards. In particular, this thesis aims at evaluating the performance of one of the most prominent LPWAN technology, LoRaWAN™, in a typical urban scenario.

Comprehensive and accurate system-level simulations of LoRa networks that consider several end-nodes that are deployed in a realistic propagation scenario, with streets and buildings, are still missing. In order to accurately simulate a LoRaWAN network, a model for a LoRa network is first proposed and then implemented to develop a new module in one of the most accurate system-level network simulators that are currently available: ns–3. Different simulations are then performed with this new tool in order to evaluate throughput, coverage, and many other important metrics that can be used to design an efficient network employing LoRa technology.

1.3 Goals

This master's thesis presents an optimal LoRa gateway placement based on Fuzzy C-Means for IoT applications, called PLACE. We introduced an algorithm for LoRa gateway placement, namely, Delay, coverage, CAPEX, and OPEX. PLACE considers two steps for LoRa gateway placement. Initially, PLACE determine the number of clusters, which means the exact amount of LoRa gateway using the Gap statistics method. Then the number of clusters means the precise amount of LoRa gateway, which is input for determining the gateway placement using the Gap statistics method.

Thus, the objectives of this work include:

- Study the essential elements of LPWAN.
- Study the essential elements of Gateway placement.
- Evaluate the performance of the proposed placement algorithm compared to other algorithms.
- Development of a Fuzzy C-Means algorithm.
- Advances the state-of-the-art in LPWAN deployment.
- Implements and evaluates the proposed algorithm.

1.4 Text organization

The rest of the document is organized as follows:

Chapter 2 gives an in-depth description of the technologies on what the proposal is based on. Then covers the network evolution and characteristics of LoRa and LoRaWAN, besides there is a categorization of clustering algorithms, giving greater emphasis to K-means and fuzzy c-means.

Chapter 3 presents related works and state-of-the-art. It also presents proposals based on different approaches and a summary of the requirements achieved by each work.

Chapter 4 describes the proposal of this master thesis and its complete analysis. It starts from the designed architecture following by showing the propagation model that was used. After that, we demonstrate each step to achieve the positioning results.

Chapter 5 presents the evaluation methodology, the achieved results, and a discussion about them. The simulation parameters are detailed, as well as some interesting behaviors are presented.

Chapter 6 there is a double objective. On the one hand, the present work intends to provide an overview of it, in what it contributed, and the conclusions resulting from its realization. On the other hand, we want to illustrate what can still be done so that more findings can be generated from what has already been presented throughout this master thesis.

CHAPTER 2

Theoretical Reference

This chapter presents LPWAN, which emerges to connect devices that require long-range and low-cost (bandwidth and power) communication services, as expected in many IoT application verticals (*e.g.,* smart grid, smart metering, smart city, smart home, and others). In light of this, LoRaWAN is considered the most adopted LPWAN technology by enabling flexible long-range communication with low power consumption and low-cost design perspectives. One of those protocols is LoRa, developed by Semtech and the LoRa-Alliance [4]. It can be said that LoRaWAN technology consists of two parts, LoRaWAN and LoRa, as shown in Figure 1. The former is a network architecture, and the latter is a protocol for the physical layer. In addition to contextualizing the research area of inductive machine learning, in which this research project is inserted. Therefore, it presents an overview of the area, discusses some of its main objectives, and briefly presents some relevant concepts involved.

2.1 LoRa Overview

LoRa is a proprietary spread spectrum modulation based on CSS by Semtech [8]. Unlike other wireless systems that use Frequency Shifting Keying (FSK) modulation for low power consumption, LoRa is based on CSS modulation. LoRa maintains the same low power characteristics as FSK modulation but increases the range of communication significantly. CSS was first used to provide military communication, where LoRa is the first low-cost implementation for commercial use [65].

LoRa is the physical layer protocol. The LoRa protocol is a proprietary protocol developed by Semtech, unlike the LoRaWAN protocol, which is open source. Due to LoRa being a proprietary protocol, information about the design and implementation is not readily available from Semtech. However, the implementation of the protocol is

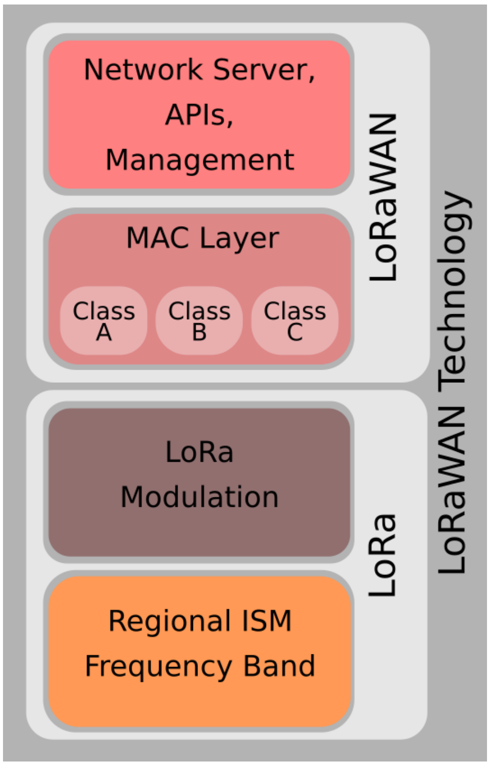

Figure 1: Protocol layers
Source: Author

considered well understood.

2.1.1 LoRa Modulation

LoRa utilizes a spread spectrum technique called CSS that was initially developed for radar applications in the 1940s [56]. In LoRa, the spreading of the spectrum is achieved by generating a chirp signal that continuously varies in frequency [56].

The main concept behind CSS is that a sinusoidal signal of linearly varying frequency and determined duration, called chirp, can be employed to "spread" information over a broader spectrum than it would typically need to occupy. This uniform distribution of a symbol over a larger bandwidth provides resistance to frequency-selective noise and interferers, at the price of lower spectral efficiency. Using some additional precautions, CSS can also be more resilient to multi-path interference and the Doppler effect than other more conventional modulations.

The Figure 2 illustrates different types of chirps - the first half being standard up chirps where frequency increases over time and restarts from the min frequency (f_{min}) towards max frequency (f_{max}) followed by short down chirps annotated as Start Frame Delimiter (SFD) that goes from fmax to fmin, and then modulated chirps that contain data bits.

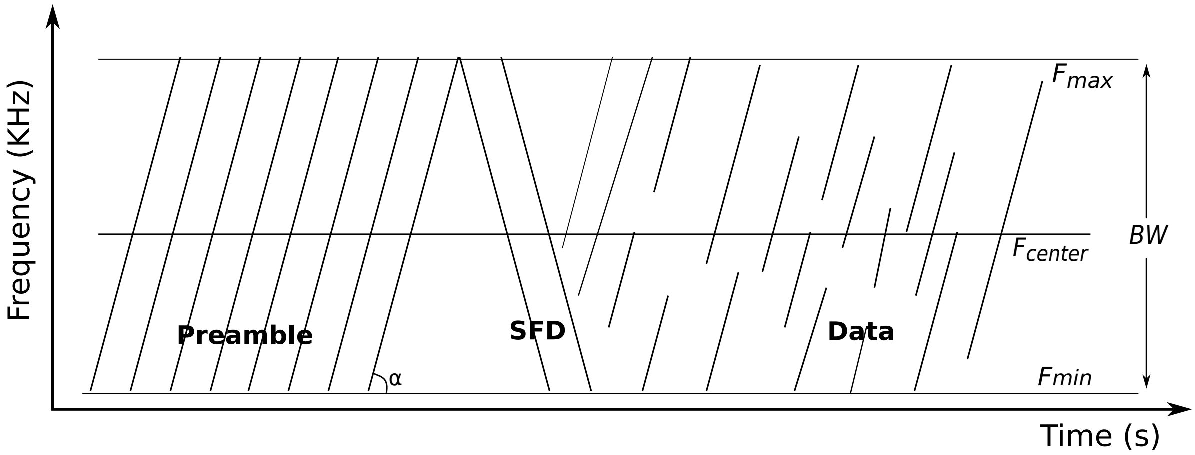

Figure 2: Spectrogram representation of a LoRa signal
Source: Adapted from Haxhibeqiri et al.

Figure 2 also shows an illustration of a LoRa packet, which presents a spectrogram representation with time on the horizontal axis and frequency in the vertical axis. Notably, a PHY layer LoRa message consists of the chirp signal sweeping the frequency band. After some repetitions of this frequency sweep that constitute a preamble (whose minimum length is of 4.25 chirps), data is encoded in the signal as instantaneous changes in the frequency of the chirp, or lack thereof. The SF is the number of bits that LoRa encodes in a symbol is a tunable parameter can be defined in Equation 2.1, where R_s is the symbol rate and R_c is the chip rate.

$$SF = \log_2 \left(\frac{R_c}{R_s} \right) \tag{2.1}$$

This means that a chirp using spreading factor SF represents 2^{SF} bits using a symbol and that there are M = 2^{SF} possible starting frequencies for a chirp. A transmissions spreading factor is also used to determine the duration of a symbol (T_s), according to the following expression:

$$T_s = \frac{2^{SF}}{BW} \tag{2.2}$$

This implies that assuming the modulation is using a fixed bandwidth, an increase of the spreading factor of 1 will yield symbols that last twice the duration. Analogously, a more significant bandwidth increases the rate at which chirps are transmitted, and consequently, the bitrate of the modulation. An increase in the transmit time for a chirp (*i.e.,* a symbol) gives the message higher robustness to interference or noise.

On the other hand, this effect may be partially balanced by the fact that for higher SF, the number of possible symbols increases. Thus, making the occurrence of symbol errors more likely: the reason for this is that achieving synchronicity between the receiver and the signal especially critical when low data rates are employed. Another disadvantage of transmitting longer messages is the increased probability of collisions. Because of the reasons above, the choice of SF affects receiver sensitivity, which is defined as:

$$S = -174 + \log_{10}(BW) + NF + SNR \quad dB, \tag{2.3}$$

where the first term is due to thermal noise at the receiver in 1 Hz of BW, NF is the noise figure at the receiver, and SNR is the signal to noise ratio required by the underlying modulation scheme. Table 1 represents SNR values for different spreading factors, where it is possible to visualize that increasing the spreading factor allows for better sensitivity.

Table 1: SNR values for different spreading factors.

SF	**SNR**
7	-7.5 db
8	-10 db
9	-12.5 db
10	-15 db
11	-17.5 db
12	-20 db

The bitrates for a range of spreading factors and bandwidths can be found in Table 2. Given Eq. 2.2, we can now get the bitrate (R_b) for a certain pair of SF and BW using a simple computation:

$$R_b = \frac{SF}{T_s} \tag{2.4}$$

Table 2: Relation SF / BW

SF	125 kHz	250 kHz	500 kHz
7	6835	12671	27343
8	3906	7812	15625
9	2197	4396	8793
10	1220	2441	4882
11	671	1342	2685
12	366	792	1464

A SF change is also translated in a change on the Time on Air of the information sent. Lowering the SF means increasing the data rate, and it means lowering the Time on Air (ToA). If a node needs less ToA, this time is available for other nodes to transmit. If the ToA is low, it results in battery consumption savings. This capability of changing the SF, and consequently, the data rate and the Time on Air, is the Adaptive Data Rate (ADR). ADR works with symmetrical uplink and downlink conditions [4].

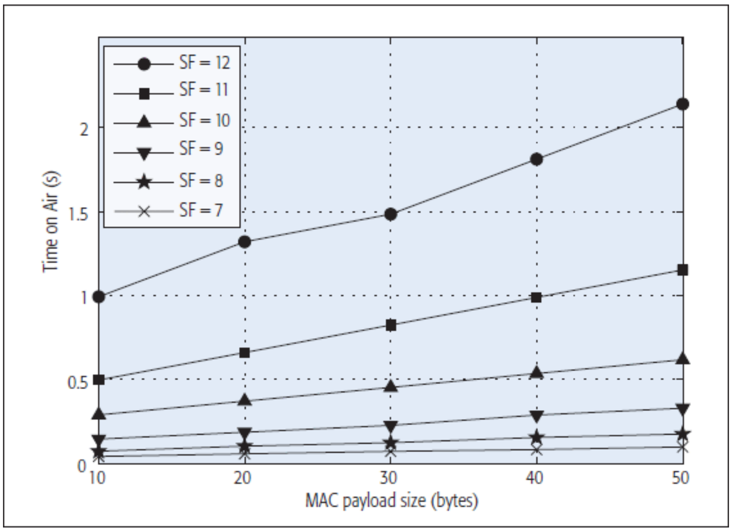

Figure 3: Time on Air of LoRaWAN with 125 kHz bandwidth.
Source: [1]

Equation 2.1 shows that as the spreading factor increases, the data rate decreases; however, at the same time, the sensitivity is higher. Table 3 shows some examples of these relations between spreading factor, channel bandwidth, bit rate, and sensitivity.

Table 3: Relation SF - BW - R_b - Sensitivity

SF	**BW**	$\boldsymbol{R_b}$	**Sensitivity**
7	125	6835	-138
8	125	3906	-136
9	125	2197	-134
10	125	1220	-131
11	125	671	-128
12	125	366	-125

2.1.2 Spreading Factor Orthogonality

One compelling feature of the LoRa modulation is that different spreading factors are pseudo-orthogonal, even when the same center frequency and bandwidth settings are used. This allows a receiver to detect a packet using spread correcting factor i even if it is overlapping in time with another transmission employing spreading factor j, as long as $i \neq j$ and the received packets Signal to Interference plus Noise Ratio (SINR) is above a certain threshold (also called isolation) that depends on both i and j. This pseudo-orthogonality between different packets leaves a network employing LoRa devices to exploit various spreading factors to achieve higher throughput concerning more traditional modulation schemes, in which a collision can cause the incorrect reception of both the intended packet and the interferer. While isolation margin is never explicitly stated in Semtech documents, in [28], this was investigated, and some estimates were made based on a model of the LoRa simulation.

2.2 LoRaWAN Overview

The LoRa-Alliance describes LoRaWAN technology [4] as:

> The LoRaWAN® specification is a Low Power Wide Area (LPWA) networking protocol designed to wirelessly connect battery operated "things" to the internet in regional, national or global networks, and targets key Internet of Things (IoT) requirements such as bi-directional communication, end-to-end security, mobility and localization services.

As can be seen from the above quote, the main focus of LoRaWAN is to be a simple network protocol that is easy to deploy and fulfills all the basic requirements for wireless battery operated IoT devices.

2.2.1 Network Topology

LoRaWAN is a technology developed by a non-profit association named LoRa Alliance, where on top of the LoRa physical layer, LoRa Alliance defined the higher layers and network architecture of LoRaWAN [37]. LoRa Alliance counts with 500+ associated members and 100+ LoRaWAN deployments in different countries.

The standardization effort focuses on massively deploy a low-cost ecosystem with long-lasting battery lifecycle, bi-directional communication, adaptive data rates, and security schemes [4]. LoRaWAN operates in an unlicensed spectrum, although operation in the licensed spectrum would also be possible [24].

LoRaWAN describes the communication protocol and system architecture for the network; on the other hand, the LoRa physical layer allows the long-range wireless communication link. Because of its influence on the communication protocol and network architecture, LoRaWAN is primarily responsible for a device's battery life, network capacity, quality of service, security, and the variety of applications served by the network [4].

Currently, several papers that analyze LoRaWAN performance have been published over the past few years [34, 52, 33, 49, 38, 51], which introduced the advantages and limitations of LoRaWAN when applied in different applications with different characteristics. Such as types of data transmission patterns, latency requirements, scale, and geographic dispersion, among others.

Usually, four elements in a star topology composes LoRaWAN, namely: (*i*) IoT devices, (*ii*) LoRaWAN Gateway, (*iii*) Network Server, and (*iv*) Application Server [41](as shown in Figure 4). IoT devices might be some sensor or other entity producing data that it wishes to relay to a network server. A LoRaWAN gateway receives data from one or multiple IoT devices connected to it over LoRa. It forwards it to the network server, acting as a transparent relay between the device and network server. A single device can also be connected to several gateways. The network server then makes the data available to an end-user/application. Communication between an end-device and a gateway is over the LoRa protocol (see chapter 2.1).

The communication between a gateway and a network server is over TCP/IP, meaning in some way a gateway connected to the Internet. A LoRaWAN gateway can negotiate data rates to increase spectral efficiency, battery life, and range, RF output power, and frequency-channels to use with end-devices using an adaptive data rate scheme. Furthermore, LoRaWAN supports broadcasts from gateways and bi-directional communication, although with limitations. These limitations reflect the use cases for the end-devices, resulting in three classes of end-devices. These classes are described in section 2.2.2.

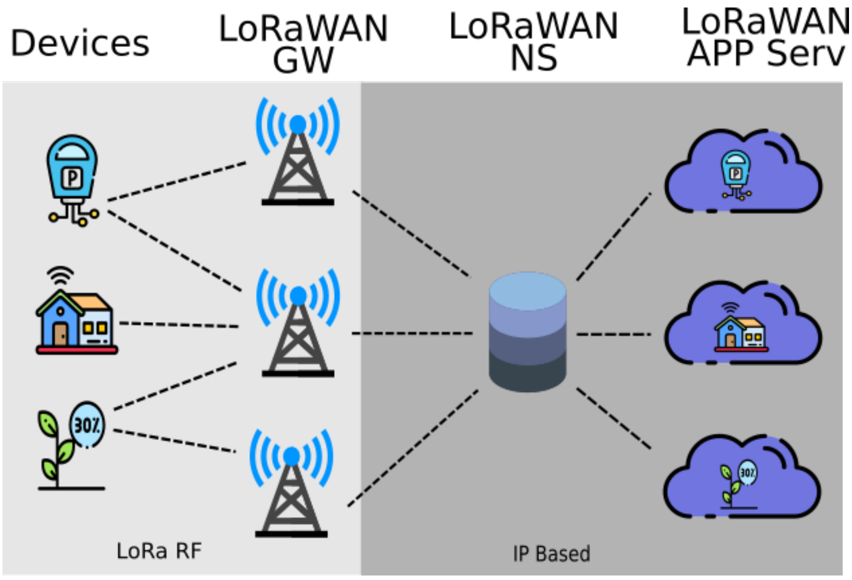

Figure 4: LoRaWAN Topology
Source: Author

In this sense, the network architecture considers a star-of-star topology, granting a single hop between the IoT device and gateway over several channels, eliminating the need to build and maintain a complex multi-hop network. The gateway communicates with the application server through an IP network [4]. Data rates vary from 0.3 kbps to 50 kbps, and security schemes implement algorithms to achieve authentication, packet integrity, and end-to-end (E2E) encryption [4].

As mentioned previously, LoRaWAN uses a star-of-stars topology. This has some advantages and disadvantages compared to a mesh-network topology as used by some other wireless sensor networks, such as ZigBee. One of the main advantages of having a star topology is that it makes it unnecessary for end-devices to listen for incoming messages and forward them, which draws a significant amount of power. Furthermore, a star-topology does not require the end-devices to contain any routing logic, resulting in simpler end-devices. However, using a star-topology has several drawbacks compared to a mesh-topology, mainly star-topologies rely on a central node, which means that, for example, a gateway failure will take several end-devices with it offline. Furthermore, a star-topology network will have no way to recover from that failure until the gateway is back up again; meanwhile, a mesh-topology network could re-route, perhaps losing some throughput but maintaining a usable network.

Figure 5 represents the protocol stack of EDs, GWs, and of the NS. While the ED and NS stacks have an application layer, gateways are only tasked with forwarding messages between the sensor (i.e., the EDs) and the NS, and are consequently totally transparent to the end application of the device, which is logically connected directly to the one on the NS.

Additionally, to the topology of the network, Sornin et al.[57] also describes the communication protocol. This includes the format of PHY and MAC layer packets, a set

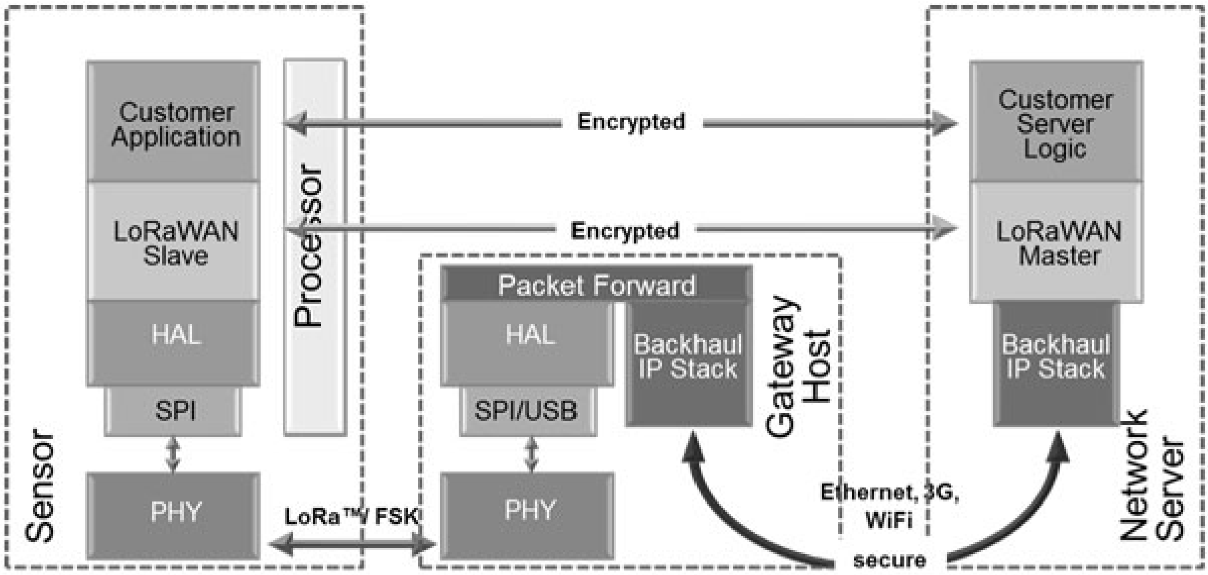

Figure 5: Protocol stacks of the various devices in a LoRaWAN.
Source: [39]

of network parameters, like the SF and channel frequencies used by an end device, and the MAC commands, which must be used to tune the settings above.

2.2.2 Devices Classes

LoRaWAN proposes three types of classes to cater for different application requirements. The three classes are class A (All end nodes), B (Beacon), C (Continuous listening) [58]. Class A is the first option that all LoRaWAN compliance end nodes should be able to support. While class B and C being mutually exclusive additional features on top of class A. This implies that class C end nodes should not implement class B and vice-versa.

Class A: Bi-directional, asynchronous, starting with an uplink message (from the device to the server) send at a scheduled uplink transmission window. Follow by opening two short receive windows by the device. If the server could not answer in either of these receive windows, the next opening will be after the future uplink transmission from the device. Only one message can respond. It has high latency and low energy consumption.

Class A (Figure 6) devices implement a node-initiated transmission where all downlink communication from the server to end nodes have to be initiated by an uplink transmission from the end-node to server. End-nodes are required to schedule two reception slots after the initial broadcast to enhance the reception of downlink communication. The first reception slot uses the exact settings as prior transmission while the second reception slot uses a preprogrammed SF. The first and second reception slots are initiated following a transmission after a T_{RX}_DELAY1 and T_{RX}_DELAY2, respectively. These reception slots only act as a preamble detection window. In any case, where the packet in the first reception window is designated for the end nodes, the second reception window would not be opened. Similarly, if the first reception window duration exceeds T_{RX}_DELAY2, the second reception window will also be aborted. LoRaWAN suggests five Tsym as timeout for each reception window.

Figure 6: Class A
Source: Adapted from Alliance

As the name suggests, class B incorporates beacons to improve the responsiveness of end devices, as shown in Figure 7. End nodes are expected to obtain T_Beacon from gateways and wake up every T_Beacon to synchronize itself with gateways to open a short reception window when necessary. This reception window will allow gateways to transmit any command from the server to end nodes within a T_Beacon. Devices from this class extend class before by adding scheduled receive windows for downlink messages. It has average energy consumption and low latency.

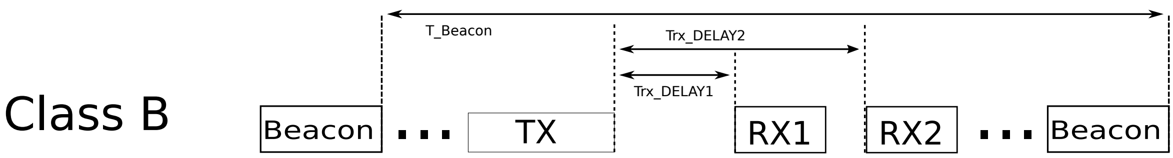

Figure 7: Class B
Source: Adapted from Alliance

Figure 8 shows class C is designed for real-time applications that require an immediate response from end nodes. With this class, end nodes are needed to open a continuous reception window using the preprogrammed SF. Instead of waiting for T_{RX}_DELAY1 before the reception, class C devices immediately open a reception window with preprogrammed SF for T_{RX}_DELAY1. After T_{RX}_DELAY1, end nodes switches the reception settings to settings used in transmission before going back to preprogrammed settings after T_{RX}_DELAY2. Devices from this class extend the class A by holding the receive windows open unless they are transmitting. It has high energy consumption, lowest latency.

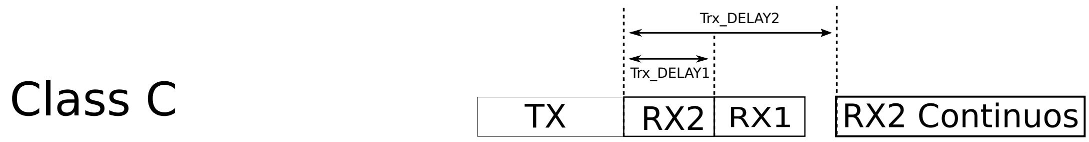

Figure 8: Class C
Source: Adapted from Alliance

LoRaWAN incorporates additional overhead into the packet to enable the MAC features provided by the classes above. These overheads include MAC headers, commands, and Message Integrity Checks (MIC). By combining the overheads, LoRaWAN imposes 13 to 27 Bytes of overhead onto each packet transmitted by an end node.

2.2.3 Frequency Bands

In [1], three regions are specified in which LoRaWANs are expected to perform at established frequencies, based on local regulations in Europe, China, and the United States as shown in Table XX. For each one of these regions, the standard mandates customized parameters that define the preamble, channel frequencies, allowed spreading factors, maximum payload size, receive windows, and Join procedures to make sure that LoRaWAN always complies with the local law.

Table 4: Frequency bands for various regions

Region	**Frequency Band [MHz]**
Europe	868–870
US	902–928
China	779-787

2.3 Clustering

The goal of clustering is that the object within a group is similar or related to one another and different from the objects in other groups [22]. Furthermore, cluster analysis explores unknown groups of data and displays a comprehensible description of the group's main similarity feature.

In this master thesis, after conducting the clustering algorithms, it hopes to

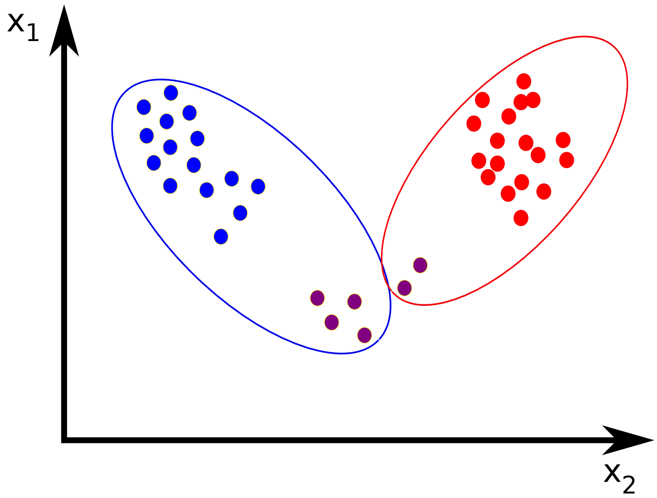

Figure 9: Illustration of K-Means
Source: Author

discover some combinations of data values automatically from the clustering results, which might predict several certain things or provide some new insights, rather than propose hypothesizes by ourselves.

2.3.1 K-Means

K-Means Clustering is a technique that performs clustering through the partitioning method [45]. This method consists of constructing multiple data partitions and evaluating them using some criteria. The name of the algorithm, K-Means, represents that "k" values declared as centers are used, and their values are represented by the means (means) of the cluster points assigned to them.

The K-Means algorithm seeks to find the center of regions that represent certain types of data. The algorithm ranges from assigning to each point to the nearest center, and choose the cluster center as the average of the points that were assigned to it. The algorithm ends when cluster assignments do not change, the moment where the convergence point of the same is reached[45], as shown in Figure 9.

Consider data whose proximity measure is Euclidean distance. The objective function, which measures the clustering corresponding to the data, uses the Sum Squares of the Error (SSE), also called dispersion. The Euclidean distance from each point to the nearest center is calculated, then the sum of squares of the residual error is calculated. The best choice for the center is the one that provides the smallest dispersion value, that is, the center that provides the best representation for the data in your cluster. The value of the sum of squares of the residual error is given by the Equation 2.5:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2 \tag{2.5}$$

Where *dist* is the Euclidean distance between two objects in a Euclidean space. The center value that minimizes SSE is the average value between the points assigned to that center. The cluster center value i^{th} is defined by the Equation 2.6:

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x \tag{2.6}$$

2.3.2 C-Means

The most popular fuzzy clustering algorithm is fuzzy c-means (FCM). It is better than K-Means (which is a hard algorithm) to avoid the local minimum, but FCM can still converge to a local minimum of squared error-based criterion. The development of membership functions is the most important problem in fuzzy clustering; Different choices include those based on similarity decomposition and cluster centroids. A generalization of FCM was proposed in Bezdek, 1981, through an objective family of functions (criterion). FCM can be taken as a generalization of the ISO-DATA algorithm.

Clustering can be classified as Soft clustering (Overlapping Clustering) and Hard Clustering (or Exclusive Clustering). In hard clustering, each object has two options, to belong or not in one cluster. Opposite, in the case of soft clustering, the objects may belong to two or more clusters with different degrees of membership. In this option, data will be associated with appropriate membership value. This means that each cluster contains memberships, and each of them is characterized by a degree value between 0 and 1.

This technique was introduced by Jim Bezdek in 1981 [6]. A basic difference between FCM and K-means is that FCM is taking more time for computation than that of K-means. The time complexity of K-mean algorithm is $O(ncdi)$ and time complexity of FCM is $O(ndc^2i)$ [27].

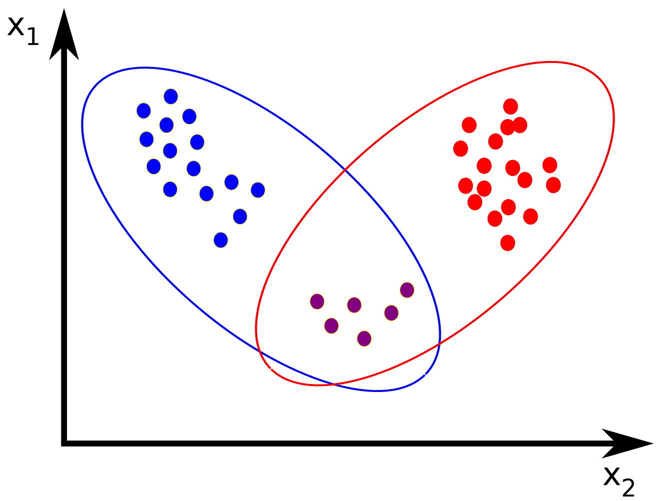

Figure 10: Illustration of Fuzzy Clustering
Source: Author

The fuzzy criterion function, *e.g.,* a weighted squared error criterion function can possible is is represented by the Equation 2.7 [35]. Which m the fuzzy factor is any real number greater than 1, μ_{ij} is the degree of membership of x_i in cluster j, x_i is the i the of d-dimensional measured data, c_j is the d-dimension center of the cluster.

$$Q = \sum_{i=1}^{N}\sum_{j=1}^{K}(\mu_{ij})^m \left\|x_i - c_j\right\|^2, \tag{2.7}$$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, and then with the update of membership μ_{ij} by Equation 2.8. Moreover, each round new cluster centers c_j are computed with the Equation 2.9.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \tag{2.8}$$

$$c_j = \frac{\sum_{i=1}^{n}(\mu_{ij})^m \cdot x_i}{\sum_{i=1}^{n}(\mu_{ij})^m} \tag{2.9}$$

This matrix stores a collection of neighborhoods that are available to all pairs of n individuals. It is represented by a $n \times n$ table. Where d(i, j) is the difference or dissimilarity measured between individuals i and j. In general, d(i, j) is a non-negative number that is close to zero when individuals i and j are very similar and become greater the difference between individuals. According can be seen by the Matrix 2.10 d(i, j) = d(j, i) and d(i, i) = 0. Dissimilarity measures will be addressed later in this chapter.

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix} \tag{2.10}$$

2.4 Final Remarks

This Chapter provided good insight into the theme of this master thesis. We Start with LoRa, explaining the definition of LoRa, what it is; then, we show how it is modulated, and to finish, we show a little bit of how the spreading factor orthogonality works. Next, we define LoRaWAN, going through its topology, the protocol stack, and finally, the definition of the device classes. Finally, we talk about clustering, starting with K-means and ending with Fuzzy C-Means. All these topics presented in this Chapter serves as a basis for a better understanding of this dissertation. Now it is time to move towards the related work from the literature.

CHAPTER 3

Related Work

This section is structured to provide an overview of the most significant contributions given by various academic papers, and categorize them according to the specific topic they were centered on: either modulation and propagation performance of LoRa devices, simulation of a LoRa system or other kinds of contributions, like protocols. Thus base station or gateway placement algorithms for IoT networks.

3.1 LoRaWAN

This subsection will mainly deal with works related to LoRa and LoRaWAN technology with works with different focuses, whether in evaluating a testbed LoRa, simulating to test the scalability of a LoRa network, simulating the traffic capacity of a network.

Sanchez-Iborra et al. [55] present a comprehensive evaluation of LoRa under different environmental conditions. The results are obtained from three real scenarios, namely, urban, suburban, and rural, considering both dynamic and static conditions, hence a discussion about the most proper LoRa physical-layer configuration for each scenario is provided. First, the estimated signal level in these scenarios has been evaluated using a precise planning tool, which employs topographic maps and the Okumura-Hata model. After that, the attained outcomes from the theoretical study have been validated by an extensive sampling campaign. In the most adverse scenarios, i.e., urban and suburban, coverage ranges around 6 km were attained; in turn, in the open scenario (rural), a long transmission distance of over 18 km with the lowest data-rate was achieved However, they use a small number of devices; furthermore, there is no stress enough to have a packet collision problem.

Pasolini et al. [47] highlight the different technologies and network topologies, even when addressing the same urban scenario and present two smart city testbeds developed in Italy. The first one concerns a smart infrastructure for public lighting and relies on a heterogeneous network using the IEEE 802.15.4 short-range communication technology, whereas the second one addresses smart-building applications and is based on the LoRa low-rate, long-range communication technology. They consider the long-range LoRa technology, and it has been experimentally shown that its maximum coverage in a dense urban environment is in the order of 12 km, which is well below the 15 km claimed by LoRa manufacturers and vendors. Remarkably, this result has been obtained in very favorable conditions, with the gateway placed at the height of 71 m above the ground and the highest possible spreading factor.

Bor et al. [9] explore the performance and capability of LoRa and LoRaWAN, then show how such a transceiver can be established to use efficiently in a wide-area application scenario. Then they demonstrate how unique features such as concurrent non-destructive transmissions and carrier detection can be employed. They conclude that LoRa radios can be used in more generic network layouts than the one used by LoRaWAN. Georgiou and Raza [25] investigate scalability simulating a single LoRaWAN gateway under interference conditions. Specifically, they propose a solution for two link outages scenarios, *i.e.,* the first when the received SNR is below the limits of the acceptable SF parameters and the second for co-spreading sequence caused by simultaneous transmission. However, a massive device's deployment likely requires multiple LoRaWAN gateways positioned close to each other, and the proximity will increase the access interference. Voigt et al. [64] investigate two alternatives for the interference problem under multiple gateway and multiple devices conditions. They conclude that the application of directional antennas and multiple gateways both improve the network performance.

Adelantado et al. [1] explain the field of LoRaWAN by investigating the limits of the technology, balancing them to application and declaring the open research challenges. In the LPWAN M2M fragmented connectivity space, there is not a single answer for all the viable connectivity requirements, and LoRaWAN is not an exception. A LoRaWAN gateway, covering a range of several Kms and able to serve up to thousands of end-devices, must be carefully dimensioned to meet the demands of each use case. Thus, the combination of the number of end-devices, the selected SFs, and the number of channels will determine if the ALOHA based access and the maximum duty-cycle regulation fit each use case. For instance, we have seen that deterministic monitoring and real-time operation cannot be guaranteed with the current LoRaWAN state of the art.

Gambiroža et al. [23] give an overview of LoRaWAN network capacity the amount of traffic that a network can handle at any given moment. The maximum capacity can be theoretically obtained under perfect synchronization and scheduling of the nodes. The aggregate capacity of a LoRaWAN gateway is evaluated as the number of end-devices that could be supported within the corresponding cell area by all six available spreading factor networks while fulfilling the QoS coverage requirement. First, they provide an overview of commonly used LPWAN solutions; NB-IoT, LTE- M, Sigfox, and LoRaWAN, and the cost, data-rate, Battery Lifetime and others aspects, considering different regions(*e.g.,* Europe, Japan, Americas). Then, a study of LoRaWAN in terms of capacity; they discuss existing solutions and highlights used approaches. Lastly, they give open challenges and opportunities.

3.2 Gateway Placement

This subsection will present some works that were necessary for the development of the positioning algorithm. Their primary focus is to improve the network through positioning. Some works solved this problem with Linear programming, others with heuristics. For this, some works have tested on testbed others with simulations.

Caillouet et al. [11] propose a theoretical framework for maximizing the LoRaWAN capacity in terms of the number of end nodes when they all have the same traffic generation process. The model optimally allocates the SF to the nodes so that attenuation and collisions are optimized. They use an accurate propagation model considering the Rayleigh channel, and we take into account physical capture and imperfect SF orthogonality while guaranteeing a given transmission success probability to each server node in the network. The imperfect SF orthogonality has an effect mainly be- tween nodes located near the gateways (using SF7) and those far away (using SF11 or 12). This affects the total number of served nodes in the network when the nodes too close to the gateway cannot be served. Numerical results show the effectiveness of our SF allocation policy. Their framework also quantifies the maximum capacity of single-cell networks and the gain induced by multiplying the gateways on the covered area. They finally evaluate the impact of physical capture and imperfect SF orthogonality on the SF allocation and network performances.

Tian et al. [59] study the placement of LPWAN gateways, when the gateways perform interference cancellation, and when the model of the residual error of interference cancellation is proportional to the power of the packet being canceled. They derived the symmetric crescent shaped regions where a GW can be placed, to enable decoding of both packets in collision sent by two SNs. Based on this conclusion, to get the minimum average contentions, which means to achieve maximum PDR. They designed two greedy algorithms to find the optimized location of GWs. One algorithm is more precise but computationally complex. The other can be made to approximate the precise one, with much lower complexity closely. However, they disconsider costs and only simulate with a small number of devices. Furthermore, there is no stress enough to have a packet collision problem.

Gravalos et al. [29] present an Integer Linear Programming (ILP) that minimizes the total cost of the network concerning the deployed devices while achieving mandatory QoS requirements. They consider a set of stationary nodes (representing facilities) placed at specific locations. Each node represents a point that generates corresponding metering data and utilizes a respective IoT end device. The communication devices differ in their transmission capabilities, which also relate to their cost. Simulation results on several topologies and traffic flow scenarios evaluate The effectiveness of the proposed ILP formulation. Nonetheless, they do not use LPWAN to validate their algorithm and use of CPLEX instead of CAPEX and OPEX that are more acceptable to the market.

Araujo et al. [5] join planning of radio (i.e., SCs) and transport resources (i.e., point-to-point fiber links) using heuristics. They compare and examine to determine the advantages and disadvantages of each approach and in some cases. They showed a significant reduction in the total cost and, to a great extent, relied on the user distribution and position of the Fiber Access Point.

Rady et al. [53] classify Gateways deployment problem into two different categories: network-aware and network-agnostic. The main difference between these two is precise knowledge of end devices position. In either category, they try to answer two design questions, that is:

1) where to place Gateways, *i.e.,* to maximize received signal strength and

2) given received signal strength which GW should the ED be associated with to balance the network load.

However, the author has a fixed number of gateways, and often, several gateways are running under performance. Moreover, they do not limit the number of users per gateway. It may happen that you have a huge number of lost packets due to the lack of available channels for reading data.

Ousat and Ghaderi [46] consider the problem of planning LoRa networks in terms of gateway placement and IoT device configuration and use mixed-integer non-linear optimization for model the plan and deploys LoRa networks. The simulations use only small networks due to complexity. They develop an approximate algorithm for planning large-scale LoRa networks efficiently and compare them with the ADR algorithm. Simulation results show averaging improvements of 15% and 20% in the throughput and energy efficiency of the network, respectively.

Hossain et al. [32] bring out the calculation method from scenario assumption, network dimensioning to cost structure calculation, which is one main contribution. In the scenario assumption part, the two dimensions, coverage, and capacity are used to divide the scenarios. Based on the assumed scenarios, dimensioning is carried out also from these two aspects. The network is required to meet all the demands. The segments of the cost structure states of the cost of deploying IoT network, the calculation method is also introduced. By using the technique, knowing the input, they can get the output. If they can get more accurate data, the more precise performance analysis can be delivered.

Petrić et al. [50] describe our experimental LoRa setup in the city of Rennes - LoRa FABIAN, and they designed, performed, and analyzed measurements for it. LoRa technology offers excellent outdoor coverage either in an urban or rural area. The antenna location and especially its elevation plays a significant role in the network performances. In the best conditions, the frame losses are shallow (about 3%). One of the goals of our study was to define criteria to switch from one spreading factor to another one to guaranty the best trade-off between channel utilization and error rate. It appears that the Received Signal Strength Indicator (RSSI) alone may not be a useful metric since measurement doesn't exhibit a strong correlation. SNR could be a better candidate. A next step will be to combine uplink and downlink traffic to find if some correlation exists between measurements. This helped determine which LoRa station will be the best to join a node.

3.3 Final Remarks

Based on our analysis of the state-of-the-art, we conclude that recent studies described proposals for gateway placement, but disconsidering cost, limiting the number of devices, LPWAN parameters, even fixing a number of gateways for clustering. However, to the best of our knowledge, all of these critical features have been provided in a unified gateway placement.

Table 5: Resource Allocation Protocols Summary

Proposal	**Addressed Problem**	**Environment**	**Strategy**	**Evaluation Metrics**
Sanchez-Iborra et al.[55]	Coverage	Testbed	SF	RSSI, SNR
Pasolini et al.[47]	Capacity	Testbed	SF	DER, Collisions
Bor et al.[9]	Coverage	Testbed	SF	RSSI, SNR
Adelantado et al.[1]	Capacity	Testbed	SF	DER, Collisions
Gambiroža et al.[23]	Coverage	Testbed	SF	RSSI, SNR
Caillouet et al.[11]	SF Allocation	Simulation	ILP,SF	DER
Tian et al.[59]	GW Placement	Simulation	WBL & PGL	PDR
Gravalos et al.[29]	GW Placement	Simulation	ILP	QoS,Cost
Araujo et al.[5]	HetNets Deployment	Simulaton	Predefined Location and Users Location	JFI, QoS, Cost
Rady et al.[53]	GW Placement	Simulation	K-Means	QoS
Ousat and Ghaderi [46]	GW Placement	Simulation	MINLP	PDR, Energy
Hossain et al.[32]	Cost structure	Simulation	Heuristic Dimensioning	CAPEX, OPEX
Petrić et al.[50]	Performance Measure-ments	Testbed	LoRa FABIAN	RSSI, PER, QoS

CHAPTER 4

PLACE

In this section, we introduce PLACE, which divides the scenario into clusters, *i.e.,* the gateway radio range, using the Fuzzy C-Means algorithm. As a result, the number of cluster means the exact amount of LoRa gateway, which is the input for determining the gateway placement using the Gap statistics method. PLACE has the advantage of improving the receiving packets by overlapping the radius of the antenna and knowing the correct number of gateways, uses the Gap Statistics. We also introduced a set of heuristics for LoRa gateway placement, namely, QoS, coverage, CAPEX, and OPEX.

4.1 Network and System model

We assume two types of LoRa devices: gateways and a set of IoT devices. IoT devices are placed at random locations without mobility to collect environment conditions, such as smart metering, and send such data to a gateway. On the other hand, LoRa gateways can be placed at a given location based on a placement algorithm. We aim to optimize the LoRA gateways placement in order to minimize the CAPEX and OPEX expenditures while respecting predefined QoS requirements.

We can view the LoRa gateways placement dilemma as a cluster/set assignment problem, where number of clusters is equals to the number of available gateways, in other words, we seek to partition devices into disjoint sets such that every set can be best served by one gateway. Our proposal considers two phases (*i.e.,* processing and validation) to compute the optimal gateway placement, as shown in Figure 11.

The planning of communication systems based on LoRaWAN requires a design methodology similar to that used in cellular systems, based on the need to estimate the coverage radius in a cell through the characteristics of the server, the device, and the

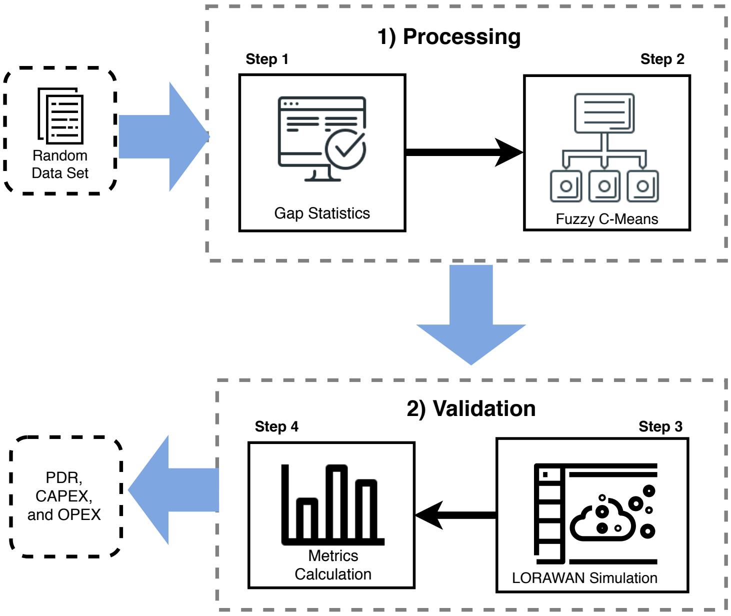

Figure 11: PLACE Overview
source: author

environment, for these situations, the prediction of the coverage area is made through mathematical models that describe the signal attenuation (path loss) for a given separation distance between the transmitter and receiver, these mathematical models are called propagation models.

Empirical propagation models are based on making several measurements and observations in real propagation environments, the equation that dictates a practical model is created in such a way as to best fit the measured data, for an empirical model to be able to represent propagation losses in a given environment efficiently, it must have its parameters derived from characteristics of the studied location, linked to the frequency of system operation and effective antenna heights used for signal transmission and reception. Below, some of the most well-known models in the literature will be addressed, with some of comparative performance analysis about the model generated in this work.

The Okumura-Hata model is well known and used in planning cellular networks, being one of the main references for projects in this area, this model was generated from graphs with information on the lost path obtained by Okumura in several measurements in the city of Tokyo, in the bands between 150 mhz and 1500 mhz. This model is valid for base stations (transmitter) with effective heights between 30 m and 200 m for customer (receiver) heights between 1 m and 10 m. In addition to being well used for mobile networks, studies show the effectiveness of this model for a LoRaWAN network [19]. The propagation models are necessary to implement on the simulation to estimate the received power by the receiver.

The following Equation 4.1 expresses the propagation loss in db units for urban areas. where: l_u is the path loss in urban areas, unit: decibel (db); h_{rx} is the height of gateway antenna, unit: meter (m); h_{tx} is the height of device antenna, unit: m; f is the frequency of transmission, unit: megahertz (mhz); c_h is the antenna height correction factor; d is the distance between the base and the device, unit: kilometer (km).

$$l_u = 69.55 + 26.16 \log_{10}(f) - 13.82 \log_{10}(h_{tx}) - c_h + [44.9 - 6.55 \log_{10}(h_{tx})] \log_{10}(d) \quad (4.1)$$

According to the adjustments made by Hata, this model can be applied to urban and rural areas using the variations of the term c_h and the adjustments shown in Equations 4.2 and 4.3. For small or medium-sized city,

$$c_h = 0.8 + (1.1 \log_{10}(f) - 0.7)h_m - 1.56 \log_{10}(f) \quad (4.2)$$

and for large cities,

$$c_h = \begin{cases} 8.29 \left(\log_{10}(1.54h_{rx})\right)^2 - 1.1 & ,if \quad 150 \leq f \leq 200 \\ 3.2 \left(\log_{10}(11.75h_{rx})\right)^2 - 4.97 & if \quad 200 < f \leq 1500 \end{cases} \quad (4.3)$$

To calculate the total loss of a path in a suburban area, the Equation 4.4 is used:

$$l = l_u - 2\left(\log_{10}\left(\frac{f}{28}\right)\right)^2 - 5.4 \quad (4.4)$$

The total loss to rural areas is found using the Equation 4.5 :

$$l = l_u - 4.78 \left(\log_{10}(f)\right)^2 + 18.33 \log_{10}(f) - 40.94 \quad (4.5)$$

4.2 Gateway Placement

We consider a partition-based clustering algorithm to allocate gateway antennas, which split the instances of the IoT devices in determined classes based on their similarity. We use this classification methodology because that instead of hierarchical clustering, which when the cluster is formed is improbable to move the devices to another group, the method used modifies the positioning of gateways and the membership of each device in each iteration of the algorithm until converging in a better organization, making it possible to find a positioning of higher packet delivery rate in relation to the first one.

4.2.1 Gap Statistics

The partition-based clustering algorithm must have the IoT devices for classifying and the number of clusters as input. Firstly, we considered the gap statistics method [60] to know the optimal number of cluster, *i.e.,* number of LoRa gateways, comparing the intra-cluster compactness of the original data features and other random data set, which means that the group number which the organization of original data farthest from random data is the most advantageous choice of clusters number.

The input is a random data set, and the implementation is defined by the following stages. Iterate over cluster numbers ($c \in [1, 2, 3..., C]$) computing the Fuzzy C-Means objective function (Jm_c) based on Equation (4.6), which gives a measure of the compactness of our clustering. It considers the cluster index c, object index i, membership coefficient μ_{ic}, fuzzification index m to control the shape of membership functions, and Euclidean distance D_{ic}^2 between the ith object and the cth cluster center. Parallel to this, we consider a predefined number of random data sets B, and for each data set ($b \in [1, 2, 3..., B]$) its computed the Fuzzy C-Means objective function ($Jm^*_{c,b}$).

$$\mathrm{Jm} = \sum_{c=1}^{C} \sum_{i=1}^{N} (\mu_{ic})^m \, D_{ic}^2 \tag{4.6}$$

The Gap statistics ($Gap(c)$) function compares the objective function computed using the original data set (Jm_c), and the objective function computed based on another random distribution of the data with the same shape ($Jm^*_{c,b}$). That returns the information of how organized the data are for each cluster number c, compared to a disorganized data set, which is computed based on Equation (4.7). In light of this, the clustered index, which maximizes the value of this function, should give a good approximation of the cluster number to be used, as shown in the Figure 12.

$$\mathrm{Gap}(c) = (1/B) \sum_b \log\left(Jm^*_{c,b}\right) - \log\left(Jm_c\right) \tag{4.7}$$

Because it is not an exact method, Gap statistics ($Gap(c)$) function usually returns an average gap value. For an accurate view of the variation of this value, the standard deviation ($sd(c)$) for each cluster number c is computed based on Equation (4.8).

$$sd(c) = \sqrt{\frac{\sum_b \left(\log\left(Jm_{c,b}\right) - \frac{1}{B}\sum_b \log\left(Jm^*_{c,b}\right)\right)^2}{B}} \tag{4.8}$$

From the standard deviation values based on Equation (4.9), we estimate the simulation error to prove the accuracy of the cluster number choice calculated by the Gap Statistics ($Gap(c)$) function. Through it, the clustered index c in ($Gap(c)$) which the growth rate starts to slow, *i.e.,* the criterion in Eq (4.10) is matched, can perform another approximation to the optimal cluster number. In addition, it is possible to calculate the Simulation Error function, as shown in the Figure 13, which the values that maximize the derived from the Equation 4.10 the values that best suit this criterion.

$$s(c) = \mathrm{sd}_c \sqrt{(1 + 1/B)} \tag{4.9}$$

$$Gap(c) \geq Gap(c+1) - s_{(c+1)} \tag{4.10}$$

Afterward, we compute the intersection of maximum values of Gap Statistic function and the Error simulation function, using both results to improve the precision of the choice of the cluster numbers, and from that, to have a cost-saving without having a significant loss of performance is choose the smallest c from this intersection to serve as input to a clustering algorithm.

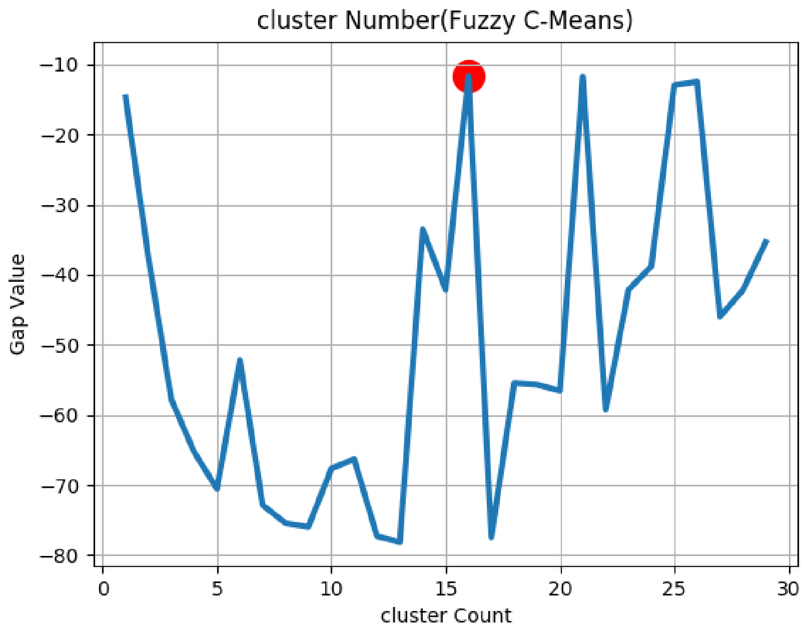

Figure 12: Gap Statistics Function
Source: Author

4.2.2 FCM

The algorithm used, Fuzzy C-Means, has a time complexity $O(ndc^2i)$ and is classified as soft clustering for being based on fuzzy logic. This heuristic, instead of hard clustering methods as K-Means has as addition the fuzziness coefficient which defines the probability of association in the range $[0, 1]$ [30]. The reason which we choose this method is that it has the advantage to classify each device with a fractional degree of membership of each gateway signal radius, making a device belongs to two or more adjacent transmission clusters (overlapping). Thus, for being in the range of n gateways, when the end-devices loses connection with one, it keeps sending message to others, significantly reducing data loss.

Secondly, PLACE considers a random data set and the number of clusters c as input, where each element has a membership percentage μ_{ic} of each cluster following the Equation (4.10), which considers data object index i and cluster index c. Afterward,

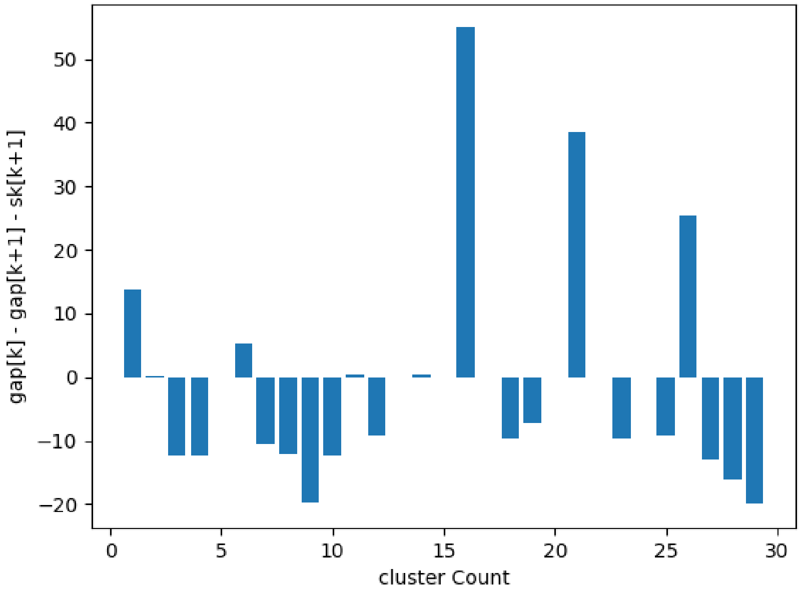

Figure 13: Gap Statistics Error Simulation Function
Source: Author

all the association of each instance we compute for each group in the c-partition matrix U, which is updated using Euclidean Distance until it reaches classification convergence. Besides that, the algorithm returns, as a result for the LoRa simulator, the exact position of the center cluster, which, together with original objects data set make up the position scenario for simulation.

$$\sum_{c=1}^{C} \mu_{ic} = 1, \qquad \text{for all } i = 1, 2, \ldots, N \tag{4.11}$$

Matrix 4.12 is a representation of a Hessian matrix. This is diagonal and all of its terms are positive. This allows us to conclude that it is positively defined.

$$U^r = \begin{bmatrix} 0.0028078597022686996 & 0.01853675875426663 & ... & 0.80810103187119 \\ 0.0028078597022686996 & 0.7775588949488006 & ... & 0.015991580677400807 \\ ... & ... & ... & ... \\ 0.9663056835727758234 & 0.01853675875426663 & ... & 0.015991580677400807 \end{bmatrix} \tag{4.12}$$

As a stopping criterion for our algorithm, we use this last Equation 4.13. If the module of the difference between the two Hessian matrices is less than or equal to the expected error,*i.e.,* the matrix stops changing substantially over the algorithm iterations, the best positions for the gateways were found.

$$\left\| \mathrm{U}^{(r+1)} - \mathrm{U}^{(r)} \right\| \leq \varepsilon(\text{ tolerance level }) \tag{4.13}$$

Algorithm 1: Fuzzy C-Means

input : Data to be clustered, fuzzification coefficient m, C desired number of clusters, ϵ error stop criterion.

output: Final fuzzy c-partitioned matrix U, cluster centers $cntr$, final Euclidean distance matrix D, objective function Jm.

initialization;
Start U matrix with random values between 0 and 1;
while *convergence criterion not reached* **do**
    instructions;
    for $c \in \{1, \ldots, C\}$ **do**
        Calculate cluster centers based on Equation 2.9;
        for $i \in \{1, \ldots, N\}$ **do**
            Calculate Euclidean distance D_{ic}^2;
            Calculate objective function Jm_{ci} base on Equation 4.6;
            Update U_{ci} matrix using the membership coefficient based on Equation 2.8;
        end
    end
    if *the convergence criterion in Equation* (4.13) *is satisfied* **then**
        Stop algorithm;
    end
end

4.2.3 LoRaWAN Simulation

With the gateways positions, we need to validate the optimum position previously computed in a LoRa environment. In other words, it is required to include the basic requirements for a LoRa network to work properly, such as SF, CR, frequency or even the number of channels, which are some examples of the peculiarities of a LoRaWAN. The LoRa gateway location must be imported from previously step and evaluated several times each algorithm to get a number of packets sent, packets received, amount of packets lost duet o interference, and the number of a lost packet by no more channel available.

In Section 2.2 and 2.1, we show some definitions for the proper functioning of a LoRaWAN. The simulator must have at least three parts, one for the Final Device, one for the Channel, and finally, one for the Gateway. The first is where you configure the Class and its behavior. The Channel is responsible for implementing shadowing, path loss. Finally, to develop the part of the Gateway, it is necessary to take into consideration all the requirements of the devices and the environment and add elements of the antenna. Figure 14 shows what a simulator must have to be able to represent the network correctly.

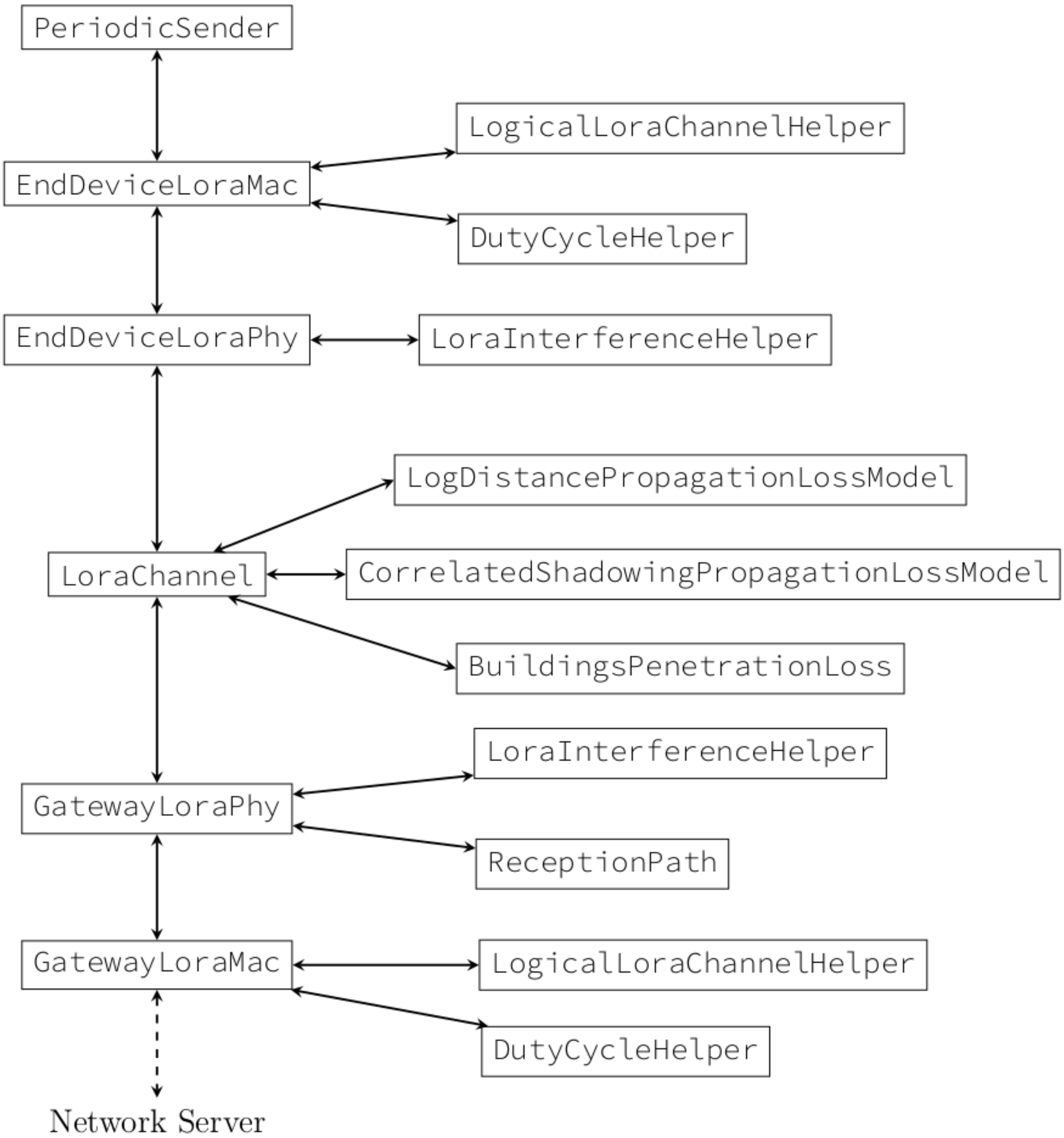

Figure 14: LoRaWAN stack
Source: [39]

4.2.4 Metrics Computation

To evaluate the performance of the proposal, we consider some metrics. To assess the performance of the network, we use the Packet Delivery Ratio (PDR). So we can compare performance with a QoS view; after all, any IoT application will require a minimum package delivery rate. To calculate costs, we use CAPEX and OPEX, based on some metrics based on project implementation work using LoRa technology.

4.3 Chapter Conclusions

PLACE receives positions from the dynamically allocated devices in order to do the clustering, but for this, it is necessary to know the amount of cluster required. So the GAP algorithm is needed. Already clustered, we use a simulator to simulate LoRaWAN and to compute the performance of our model compared to other positioning methods. The following chapter evaluates the proposal with the state of the art, and shows the improvements obtained.

CHAPTER 5

Evaluation

This section describes the methodology and metrics used to evaluate the PLACE algorithm in terms of QoS, CAPEX, and OPEX compared to random and grid deployment with different numbers of LoRa Gateway. Section 5.1 talks about the metrics and methodologies that were used to formulate the work. As an example, we explain how we got to the CAPEX and OPEX formula. In section 5.2, we show the results. Finally, in section 5.3 we talk about our conclusions about the results obtained in the previous section.

5.1 Methodology

We use python to compute the number of LoRa gateways, *i.e.,* Steps 1, and 2. Afterward, the number and location of LoRa Gateway are implemented in NS-3, which implements the LoRa protocol stack for communication between the LoRa devices and Gateway. NS-3 also implements an error model for LoRa modulation based on baseband simulations of a LoRa transceiver over an additive white Gaussian noise (AWGN) channel [61]. This allows us to reproduce a device behavior in LoRa networks, and the code is available in our GitLab page[1]. We conducted 33 simulations with different randomly generated seeds fed to the simulator's pseudo-random number generator (MRG32k3a). Results show the values with a confidence interval of 95%. To compute the costs we take into account the dissertation of Lin[36], with this information, we created Table 6 .

[1]https://gitlab.com/gercomlacis/cea/lorawan

Table 6: Cost Assumptions Equipment

Costs Description	LoRaWAN
Equipment Cost ($K€$)	1
Installation Cost ($K€$)	2
Network Setup Cost ($K€$)	0.1
Spectrum Cost ($K€/kHz/site$)	0
Transmission Installation Cost($K€$)	4
Transmission Cost($K€$)	0.1
Site Lease ($K€/year$)	1
Electricity Cost ($K€/year$)	1
Transmission Cost ($K€/year$)	0.1

We consider a LoRa Class A network, where transmissions are always initiated by the IoT devices, in a non-synchronous aspect. For this purpose, the IoT devices may choose at random one channel. One of the parameters of the system is the reporting periodicity τ. In our scenario, we assign every IoT devices a random initial reporting delay, after which the node generates a new packet every τ seconds. In this work, downlink transmissions, *i.e.,* messages from the gateways to the IoT devices, are considered. We do not consider it a significant limitation since we expect most of the traffic in a LPWAN to be uplink.

LoRa imposes the use of at least three mandatory channels at center frequencies 868.1, 868.3, and 868.5MHz in the European sub-band, which we are using. When sending, the IoT devices picks one of these three channels at random. We consider that all the gateways are transmitting with maximum power P_{max}, with an antenna gain of G_t. We deployed 1000 fixed IoT devices in a $100Km^2$ area. Table 7 shows the main simulation parameters.

Table 7: Simulation parameters

Parameters	Value
IoT Devices	1000
Simulation Area	100 Km^2
Data Message Size	20 Bytes
Time Simulation	600 s
Gateway Radius	2000m
Frequency	868 MHz
Number Channels	3
Propagation Model	Okumura Hata

We evaluate the impact of the PLACE with different placement strategies. For grid strategy, we divided the scenario into grids of 2 km each side (denoted as Grid25 in the plot), and also grids of 2.5 Km (denoted as Grid16 in the plot), where the gateway is placed at the center of each cell. For the Random, we deployed 25 gateways (denoted as Rand25 in the plot) and 16 gateways (denoted as Rand16 in the plot) randomly placed. Finally, PLACE, we compute the number of gateways using gap statistics and then find the gateways location using Fuzzy C-Means. We evaluate the investment in terms of CAPEX (Eq. 5.1), OPEX (Eq. 5.3), and total cost for the period of 1 year. We also compute the performance of LoRa network in terms of PDR.

CAPEX can be computed based on Eq. (5.1). CAPEX depends on the cost for acquisition a LoRaWan gateway C_{Bs}, the cost to deploy the gateway C_{ins}, gateway setup C_{set}, and transmission installation T_{xinst}. The deployment cost C_{ins} means to pay a team to go to the site or a designed place to install a gateway. After sending a team in loco, an engineer should set up a gateway C_{set}. These cost, *i.e.,* C_{Bs}, C_{ins}, and C_{set}, are for a device and communicate with a gateway. However, there is a cost to implement gateway-cloud communication T_{xinst}. It depends on the local and the application, and it could be LTE or even fiber.

$$CAPEX = \sum_{i=1}^{S} C_{Bs} + C_{ins} + C_{set} + T_{xinst} \tag{5.1}$$

The algorithm has some constraints to be a valid proposal, for example, the minimum percentage of covered devices (value X) in the Equation 5.2 $A_{k,i} = 1$ represents covered devices.

$$\frac{\sum_{i \in I} \sum_{k \in K} A_{k,i}}{\text{TotalUsers}} \geq X \tag{5.2}$$

OPEX is computed based on Eq. (5.3). Specifically, the Operation and Maintenance Cost C_{man}, which is about 10%-15% of the CAPEX value. Considering that all the places that the gateways will be installed will be rented, C_{lease} is concerning a lease cost. C_{elet} is regarding electricity cost per year, and C_{Trans} is relating to the transmission cost depending on the chose technology for T_{xinst}. All these costs are per year, so to predict future cost, change the variable t that represents time in years.

$$OPEX = \left(C_{man} * CAPEX + \sum_{i=1}^{S} (C_{lease} + C_{elet} + C_{Trans}) \right) * t \tag{5.3}$$

We validate using a LoRa Module [40] in a network simulator, to validate the optimum position previously computed. We import the users and gateways positions from a file, then we run 33 times each algorithm (Grid, Rand25, C-Means and Rand16) to get number of packets sent, packets received, amount of packets lost because of interference, and number of lost packet by no more channel available.We use a Grid 5x5 for a Grid algorithm, totaling 25 gateways. The same amount of gateways for the Rand25, but the gateways is place randomly. Likewise Rand16, although with 16 gateways.

5.2 Results

Figure 15 shows the number of LoRa gateways for the grid, rand, and PLACE gateway placement. We can observe that PLACE algorithm computed the mean of the number of gateway as 15.78, which is approximately 9 gateway less compared to grid and rand deployment with 25 LoRa gateways. PLACE computed the number of gateway, and their placement based on the IoT device location in order to provide high PDR, while reduces the CAPEX and OPEX, as shows the following results. Grid16 and Rand16 have 16 gateways, which is the similar number of gateways computed by PLACE in order to compare the QoS, CAPEX, and OPEX performance of PLACE, Grid16 and Rand16 with similar number of gateways.

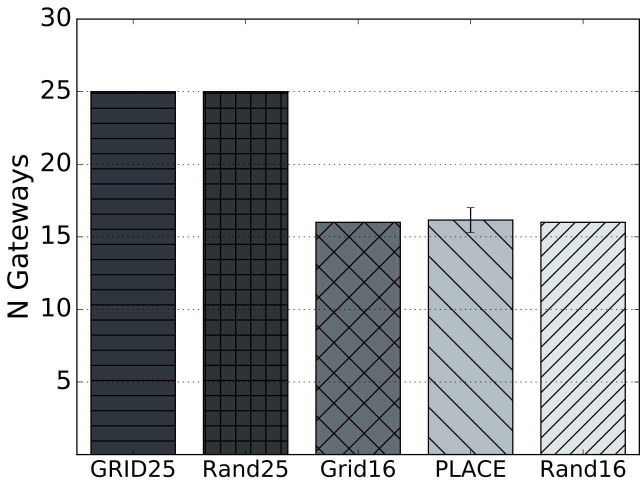

Figure 15: Number of Gateways for Different LoRa Gateway Placement Algorithms
Source: Author

Figure 16 shows the PDR for data transmitted considering grid, rand, and PLACE gateway placement algorithm. By analyzing the results we can see that 16 gateway randomly placed have worse PDR compared to algorithm with same number of gateways, *i.e.,* rand16 has 5% lower performance compared with grid16, as well as 10% compared to PLACE. This is because it randomly place the gateways without consider the location of IoT devices, and in some cases it has high concentrate place overloading the data capacity resulting packet loss. On the other hand, PLACE provides PDR similar to Rand25 besides the difference of 9 gateways. This is because GRID has more gateways to cover the entire scenario, while PLACE computed the optimum number and location of gateway based on the network IoT device location in order to provide high PDR.

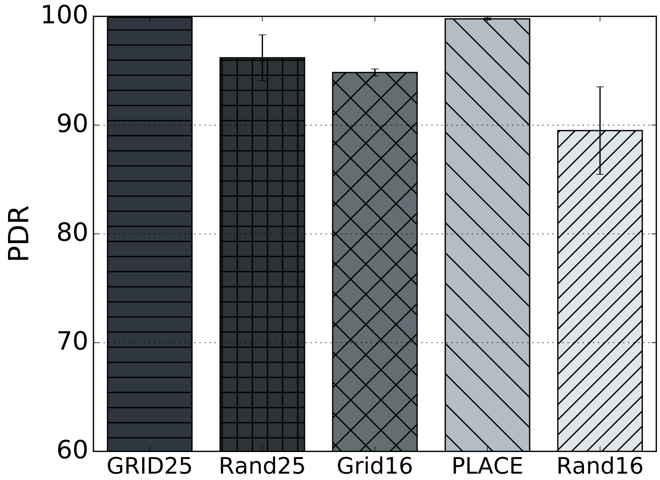

Figure 16: PDR for different LoRa gateway placement algorithms
Source: Author

Considering Figure 15, we took a seed from the simulation, the seed is the same for Figures 12 and 13. Figure 17 shows all the positions of the gateways for each one of the algorithms used to compare with PLACE. The small dots are the devices, and the blue circles are our algorithm. In this specific scenario, we use 16 gateways for our algorithm.

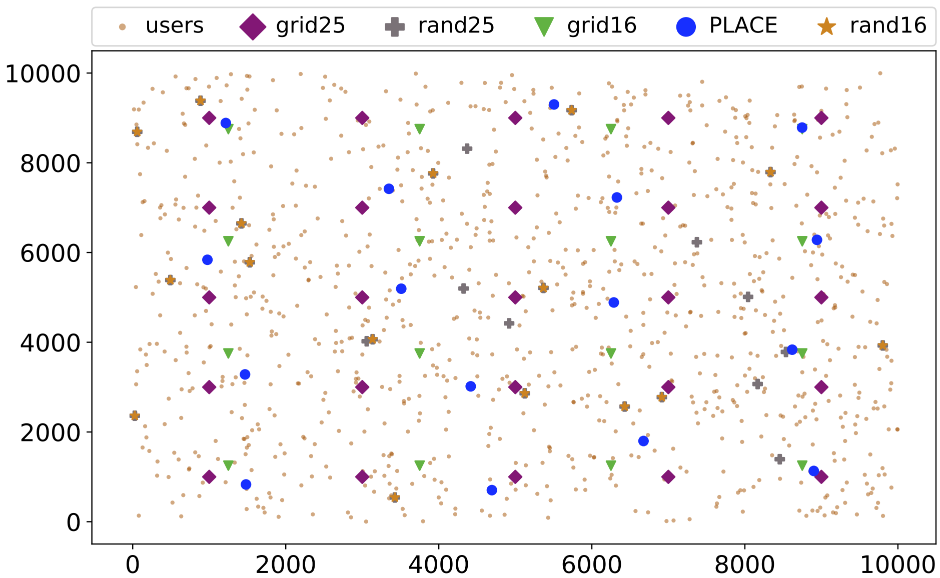

Figure 17: Number of LoRa gateways for different LoRa gateway placement algorithms
Source: Author

We considered Grid25 and PLACE, since Rand25 and Rand16 have the same number of gateways, and thus the CAPEX and OPEX will be the same. By analyzing the results of Figure 18, we can observe that the PLACE reduces in 36.36% the CAPEX, since it found the optimum number of LoRa gateway, while provides similar PDR compared to Grid25. To calculate CAPEX we use the Equation 5.1.

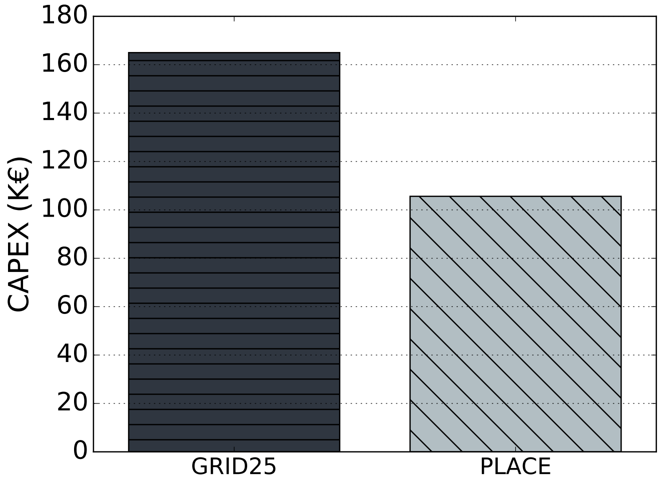

Figure 18: Total CAPEX in K€
Source: Author

Figure 19, shows that even for OPEX, the ratio between GRID25 and PLACE remains 36%. We consider $t = 1$ in Equation 5.3, but even if increasing the variable the the ratio stays the same, as there is no operation to increase or decrease in OPEX calculation. In this example we saw that PLACE costs 45024.75€ and GRID25 28815.84€. As stated earlier, the cost of the other algorithms will not change the cost, as it has the same amount of gateway.

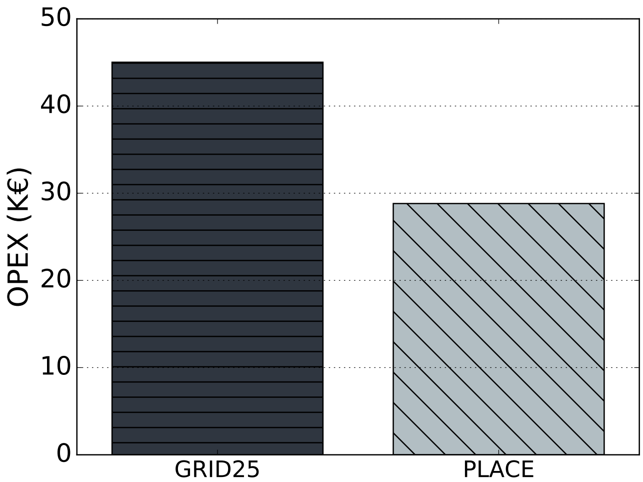

Figure 19: Total OPEX in K€
Source: Author

Moreover, Figure 20 presents a total cost for a placement in a 100 km^2 area for

1 year, where we observe that PLACE is 36% lower than using Grid. All the cost are only considering the gateways because the costs of the devices will be the same regardless algorithm we choose.

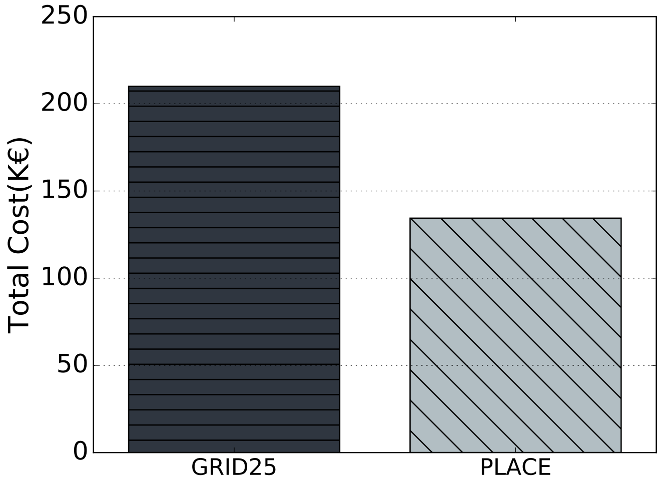

Figure 20: Total Cost(CAPEX + OPEX) in K€
Source: Author

Finally, Figure 21 shows the costs of both CAPEX and OPEX for each price associated. The cost of installing the transmission is responsible for 47% of the entire cost. This cost is related to the technology that will be responsible for sending the gateway information to a cloud (*e.g.*, LTE, WiFi, *etc*).

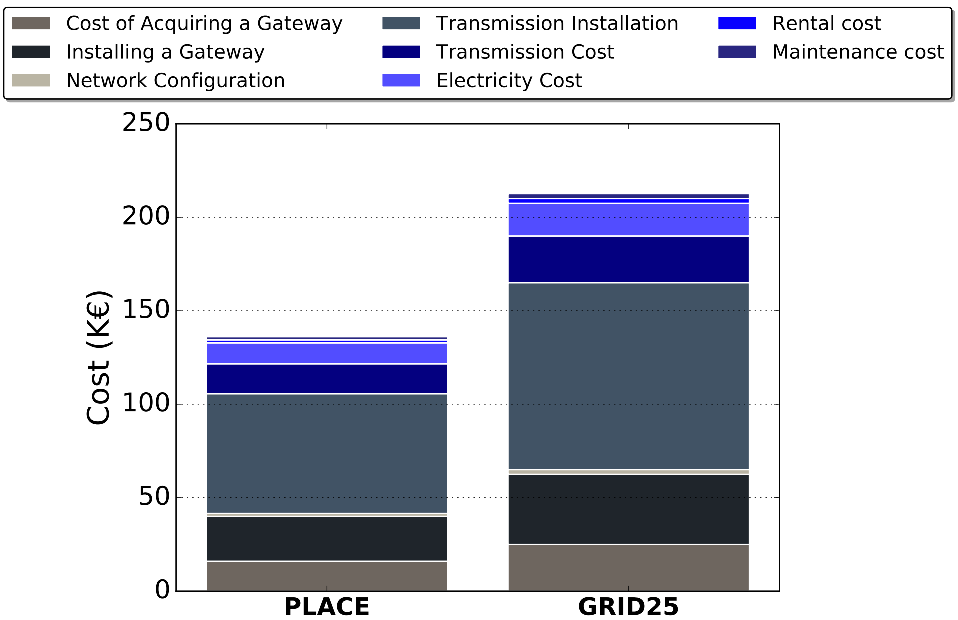

Figure 21: CAPEX and OPEX per cost
Source: Author

5.3 Chapter Conclusions

From our performance evaluation analysis, we conclude that PLACE improves in a cost-efficient way the IoT scenario, managing to maintain a high value of package delivery rate. In this sense, it is definitely possible to implement a smart grid scenario, for example, at a lower cost if you use the PLACE algorithm instead of the other algorithms compared in this chapter. In addition to using a clustering model based on Fuzzy, we can prepare the devices for possible failure in the gateways.

CHAPTER 6

Conclusions

This dissertation addresses the problem of LoRaWAN implementation in small IoT installations, such as sensor networks for agriculture monitoring and control. To solve this problem, a LoRa gateway placement algorithm is proposed. In order to reduce the overall cost of the project, maintaining the system resiliency and quality of service. Initially, the number of gateways is computed using the Fuzzy C-Means algorithm, which is the input for determining the gateway location using the Gap statistics method.

6.1 Contributions

It computes the costs of a LoRaWAN project. Since the prices are high, we use heuristics based on Fuzzy C-Means to reduce the number of the gateway and consequently to reduce the costs of implementation. Simulation results demonstrate the same Packet Delivery Rate for PLACE and GRID25 within nine gateway less.

In summary, we seek to position LoRaWAN gateway in a strategic position in order to cover all the user devices (sensors) with reliability and reducing the costs. We use Capital expenditure and Operating expenses to evaluate our algorithm. The results exhibit savings, up to 36%, on the overall installation cost, and statistically with the same delivery ratio.

In my opinion, the gateway positioning algorithm for LPWAN is an area that is still little explored, which tends to grow as it is an area that only tends to grow. As you look for the implementation of LPWAN gateways, it increases more accurate will be the algorithms to estimate the cost of the project.

6.2 Future works

For future work, the strategy is to change the clustering method to start clustering by RSSI instead of the Euclidean distance. In addition to comparing with works that implement entire Mixed Integer Linear Programming (MILP) to know how close to the optimum our result is. Another step for this work is to implement the LoRa parameters optimization with the positioning of gateways to be a self-configured and self-optimized network.

6.3 Academical Production

The results obtained in this master's thesis were published in the following events [43, 42, 41]:

1. [Matni *et al.* 2019] **Matni, N.**, Moraes J., Rosário, D., Cerqueira, E., & Neto, A. (2019, November). Optimal Gateway Placement Based on Fuzzy C-Means for Low Power Wide Area Networks. In proceedings of the IEEE Latin-American Conference on Communications (LATINCOM) (pp. 1-6). IEEE.

2. [Matni *et al.* 2020] **Matni, N.**, Moraes J., Pacheco L., Rosário, D., Nunes H., Cerqueira, E., & Neto, A. (2019, November). Experimenting Long Range Wide Area Network in an e-Health Environment: Discussion and Future Directions. In proceedings of the 16th annual International conference on Wireless Communications Mobile Computing (IWCMC 2020) (pp. 1-6). IEEE.

3. [Matni *et al.* 2019] **Matni, N.**, Rosário, D., & Neto, A. (2019). Um Ecossistema IoT para Redes Elétricas Inteligentes. Computação Brasil, 40, 41-44.

Bibliography

[1] Ferran Adelantado, Xavier Vilajosana, Pere Tuset-Peiro, Borja Martinez, Joan Melia-Segui, and Thomas Watteyne. Understanding the limits of lorawan. *IEEE Communications magazine*, 55(9):34–40, Sep. 2017.

[2] Hossein Ahmadi, Goli Arji, Leila Shahmoradi, Reza Safdari, Mehrbakhsh Nilashi, and Mojtaba Alizadeh. The application of internet of things in healthcare: a systematic literature review and classification. *Universal Access in the Information Society*, pages 1–33, 2018.

[3] Godfrey Anuga Akpakwu, Bruno J Silva, Gerhard P Hancke, and Adnan M Abu-Mahfouz. A survey on 5g networks for the internet of things: Communication technologies and challenges. *IEEE Access*, 6:3619–3647, 2018.

[4] LoRa Alliance. White paper: A technical overview of lora and lorawan. *The LoRa Alliance: San Ramon, CA, USA*, pages 7–11, 2015.

[5] Welton Araujo, Rafael Fogarolli, Marcos Seruffo, and Diego Cardoso. Deployment of small cells and a transport infrastructure concurrently for next-generation mobile access networks. *PloS one*, 13(11):e0207330, 2018.

[6] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

[7] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy. Massive machine-type communications in 5g: physical and mac-layer solutions. *IEEE Communications Magazine*, 54(9):59–65, Sep. 2016. doi: 10.1109/MCOM.2016.7565189.

[8] M. Bor and U. Roedig. Lora transmission parameter selection. In *Proceedings of the 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 27–34, June 2017. doi: 10.1109/DCOSS.2017.10.

[9] Martin C Bor, Utz Roedig, Thiemo Voigt, and Juan M Alonso. Do lora low-power wide-area networks scale? In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 59–67. ACM, 2016.

[10] Stefano Buzzi, I Chih-Lin, Thierry E Klein, H Vincent Poor, Chenyang Yang, and Alessio Zappone. A survey of energy-efficient techniques for 5g networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4):697–709, 2016.

[11] Christelle Caillouet, Martin Heusse, and Franck Rousseau. Optimal sf allocation in lorawan considering physical capture and imperfect orthogonality. In *proceedings of IEEE Global Communications Conference (GLOBECOM 2019)*, 2019.

[12] Patrik Cerwall, P Jonsson, R Möller, S Bävertoft, S Carson, and I Godor. Ericsson mobility report. *On the Pulse of the Networked Society. Hg. v. Ericsson*, 2015.

[13] Devaki Chandramouli, Rainer Liebhart, and Juho Pirskanen. *5G for the Connected World*. Wiley Online Library, 2019.

[14] Shubhajeet Chatterjee, Mohammad J Abdel-Rahman, and Allen B MacKenzie. Optimal base station deployment with downlink rate coverage probability constraint. *IEEE Wireless Communications Letters*, 7(3):340–343, 2017.

[15] Sunil Cheruvu, Anil Kumar, Ned Smith, and David M Wheeler. Connectivity technologies for iot. In *Demystifying Internet of Things Security*, pages 347–411. Springer, 2020.

[16] Alem Čolaković and Mesud Hadžialić. Internet of things (iot): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, 144:17–39, 2018.

[17] Daniele Croce, Michele Gucciardo, Stefano Mangione, Giuseppe Santaromita, and Ilenia Tinnirello. Impact of lora imperfect orthogonality: Analysis of link-level performance. *IEEE Communications Letters*, 22(4):796–799, 2018.

[18] Jonathan de Carvalho Silva, Joel J. P. C. Rodrigues, Antonio M Alberti, Petar Solic, and Andre LL Aquino. Lorawan a low power wan protocol for internet of things: A review and opportunities. In *proceedings of the 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech 2017)*, pages 1–6. IEEE, 2017.

[19] Rida El Chall, Samer Lahoud, and Melhem El Helou. Lorawan network: radio propagation models and performance evaluation in various environments in lebanon. *IEEE Internet of Things Journal*, 6(2):2366–2378, 2019.

[20] Mehmet Ali Ertürk, Muhammed Ali Aydın, Muhammet Talha Büyükakkaşlar, and Hayrettin Evirgen. A survey on lorawan architecture, protocol and technologies. *Future Internet*, 11(10):216, 2019.

[21] Dave Evans. A internet das coisas. *Como a próxima evolução da Internet está*, 2011.

[22] Adil Fahad, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y Zomaya, Sebti Foufou, and Abdelaziz Bouras. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE transactions on emerging topics in computing*, 2(3):267–279, 2014.

[23] Jelena Čulić Gambiroža, Toni Mastelić, Petar Šolić, and Mario Čagalj. Capacity in lorawan networks: Challenges and opportunities. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE, 2019.

[24] J. . Gambiroa, T. Masteli, P. oli, and M. agalj. Capacity in lorawan networks: Challenges and opportunities. In *proceedings of the 4th International Conference on Smart and Sustainable Technologies (SpliTech 2019)*, pages 1–6, June 2019.

[25] O. Georgiou and U. Raza. Low power wide area network analysis: Can lora scale? *IEEE Wireless Communications Letters*, 6(2):162–165, April 2017. doi: 10.1109/LWC.2016.2647247.

[26] Hakim Ghazzai, Taha Bouchoucha, Ahmad Alsharoa, Elias Yaacoub, Mohamed-Slim Alouini, and Tareq Y Al-Naffouri. Transmit power minimization and base station planning for high-speed trains with multiple moving relays in ofdma systems. *IEEE Transactions on Vehicular Technology*, 66(1):175–187, 2016.

[27] Soumi Ghosh and Sanjay Kumar Dubey. Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4), 2013.

[28] Claire Goursaud and Jean-Marie Gorce. Dedicated networks for IoT : PHY / MAC state of the art and challenges. *EAI endorsed transactions on Internet of Things*, October 2015. doi: 10.4108/eai.26-10-2015.150597. URL `https://hal.archives-ouvertes.fr/hal-01231221`.

[29] Ilias Gravalos, Prodromos Makris, Kostas Christodoulopoulos, and Emmanouel A Varvarigos. Efficient gateways placement for internet of things with qos constraints. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2016.

[30] Dang Thanh Hai, Trong Le Vinh, et al. Novel fuzzy clustering scheme for 3d wireless sensor networks. *Applied Soft Computing*, 54:141–149, 2017.

[31] Jetmir Haxhibeqiri, Eli De Poorter, Ingrid Moerman, and Jeroen Hoebeke. A survey of lorawan for iot: From technology to application. *Sensors*, 18(11):3995, 2018.

[32] Mohammad Istiak Hossain, Lu Lin, and Jan Markendahl. A comparative study of iot-communication systems cost structure:: Initial findings of radio access networks cost. In *2018 11th CMI International Conference: Prospects and Challenges Towards Developing a Digital Economy within the EU*, pages 49–55. IEEE, 2018.

[33] Ao Huang, Mengxing Huang, Zhentang Shao, Xu Zhang, Di Wu, and Chunjie Cao. A practical marine wireless sensor network monitoring system based on lora and mqtt. *arXiv preprint arXiv:1906.09571*, 2019.

[34] Heon Huh and Jeong Yeol Kim. Lora-based mesh network for iot applications. In *proceedings of the IEEE 5th World Forum on Internet of Things (WF-IoT 2019)*, pages 524–527. IEEE, 2019.

[35] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[36] Lu Lin. Cost structure of iot connectivity services, 2019. URL `http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-253796`.

[37] Jiangbin Lyu, Dan Yu, and Liqun Fu. Achieving max-min throughput in lora networks. *arXiv preprint arXiv:1904.12300*, 2019.

[38] M. Magno, F. A. Aoudia, M. Gautier, O. Berder, and L. Benini. Wulora: An energy efficient iot end-node for energy harvesting and heterogeneous communication. In *proceedings of the Design, Automation Test in Europe Conference Exhibition (DATE 2017)*, pages 1528–1533, March 2017. doi: 10.23919/DATE.2017.7927233.

[39] Davide Magrin. Network level performances of a lora system. 2016.

[40] Davide Magrin, Marco Centenaro, and Lorenzo Vangelista. Performance evaluation of lora networks in a smart city scenario. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2017.

[41] Nagib Matni, Jean Moraes, Denis Rosário, Eduardo Cerqueira, and Augusto Neto. Optimal gateway placement based on fuzzy c-means for low power wide area networks. In *proceedings of IEEE Latin-American Conference on Communications (LATINCOM 2019)*, pages 1–7. IEEE, November 2019.

[42] Nagib Matni, Denis Rosário, and Augusto Neto. Um ecossistema iot para redes elétricas inteligentes. *Computação Brasil*, 40:41–44, 2019.

[43] Nagib Matni, Jean Moraes, Lucas Pacheco, Denis Rosário, Helder May Nunes da Silva Oliveira, Eduardo Cerqueira, and Augusto J. Venancio Neto. Experimenting long range wide area network in an e-health environment: Discussion and future directions. In *proceedings of the 16th annual International conference on Wireless Communications Mobile Computing (IWCMC 2020)*, Limassol, Cyprus, June 2020.

[44] Renato Mota, André Riker, and Denis Rosário. Adjusting group communication in dense internet of things networks with heterogeneous energy sources. In *Anais do XI Simposio Brasileiro de Computacao Ubiqua e Pervasiva*. SBC, 2019.

[45] Andreas C Muller and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. O'Reilly Media, 2017.

[46] Behnam Ousat and Majid Ghaderi. Lora network planning: Gateway placement and device configuration. In *proceedings of the IEEE International Congress on Internet of Things (ICIOT 2019)*, pages 25–32. IEEE, 2019.

[47] Gianni Pasolini, Chiara Buratti, Luca Feltrin, Flavio Zabini, Cristina De Castro, Roberto Verdone, and Oreste Andrisano. Smart city pilot projects using lora and ieee802. 15.4 technologies. *Sensors*, 18(4):1118, 2018.

[48] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo. On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology. In *proceedings of the 14th International Conference on ITS Telecommunications (ITST 2015)*, pages 55–59, Dec 2015.

[49] Juha Petäjäjärvi, Konstantin Mikhaylov, Marko Pettissalo, Janne Janhunen, and Jari Iinatti. Performance of a low-power wide-area network based on lora technology: Doppler robustness, scalability, and coverage. *International Journal of Distributed Sensor Networks*, 13(3):1550147717699412, 2017.

[50] Tara Petrić, Mathieu Goessens, Loutfi Nuaymi, Laurent Toutain, and Alexander Pelov. Measurements, performance and analysis of lora fabian, a real-world implementation of lpwan. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2016.

[51] T. Polonelli, D. Brunelli, M. Guermandi, and L. Benini. An accurate low-cost crackmeter with lorawan communication and energy harvesting capability. In *proceedings of the IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA 2018)*, pages 671–676, Sep. 2018. doi: 10.1109/ETFA.2018.8502592.

[52] Tommaso Polonelli, Davide Brunelli, Achille Marzocchi, and Luca Benini. Slotted aloha on lorawan-design, analysis, and deployment. *Sensors*, 19(4):838, 2019.

[53] Mina Rady, Maryam Hafeez, and Syed Ali Raza Zaidi. Computational methods for network-aware and network-agnostic iot low power wide area networks (lpwan). *IEEE Internet of Things Journal*, 2019.

[54] Usman Raza, Parag Kulkarni, and Mahesh Sooriyabandara. Low power wide area networks: An overview. *IEEE Communications Surveys & Tutorials*, 19(2):855–873, 2017.

[55] Ramon Sanchez-Iborra, Jesus Sanchez-Gomez, Juan Ballesta-Viñas, Maria-Dolores Cano, and Antonio Skarmeta. Performance evaluation of lora considering scenario conditions. *Sensors*, 18(3), 2018.

[56] AN SEMTECH and Modulation Basics. An1200. 22. *LoRa Modulation Basics*, 46, 2015.

[57] N Sornin, M Luis, T Eirich, T Kramp, and O Hersent. Lorawan specification v1. 0. *0. Technical report, LoRa Alliance, Tech. Rep.*, 2015.

[58] Nicholas Sornin and A Yegin. Lorawan 1.1 specification. *LoRa Alliance*, 2017.

[59] Hongxian Tian, Mary Weitnauer, and Gedeon Nyengele. Optimized gateway placement for interference cancellation in transmit-only lpwa networks. *Sensors*, 18(11): 3884, 2018.

[60] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.

[61] Floris Van den Abeele, Jetmir Haxhibeqiri, Ingrid Moerman, and Jeroen Hoebeke. Scalability analysis of large-scale lorawan networks in ns-3. *IEEE Internet of Things Journal*, 4(6), 2017.

[62] Rob Van Der Meulen. Gartner says 6.4 billion connected 'things' will be in use in 2016, up 30 percent from 2015. *Press Release. Online at https://www.gartner.com/newsroom/id/3165317 (2018/01/15)*, 2015.

[63] Benny Vejlgaard, Mads Lauridsen, Huan Nguyen, István Z Kovács, Preben Mogensen, and Mads Sorensen. Interference impact on coverage and capacity for low power wide area iot networks. In *proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2017)*, pages 1–6. IEEE, 2017.

[64] Thiemo Voigt, Martin C. Bor, Utz Roedig, and Juan M. Alonso. Mitigating internetwork interference in lora networks. *CoRR*, abs/1611.00688, 2016.

[65] Andrew J Wixted, Peter Kinnaird, Hadi Larijani, Alan Tait, Ali Ahmadinia, and Niall Strachan. Evaluation of lora and lorawan for wireless sensor networks. In *proceedings of the IEEE SENSORS*, pages 1–3. IEEE, 2016.