LoRaWAN Simulation and Analysis for Performance Enhancement of Realistic Networks

Bruno Citoni

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

July 2022

Abstract

The Internet of Things (IoT) is becoming an ubiquitous technology, with new devices, solutions and applications being developed at an ever-increasing rate.

Fundamental to the IoT revolution is the adoption of wireless protocols purposely designed to enable low-cost, long-range communication for numerous connected devices. Low Power Wide Area Networks (LPWANs) are wide area wireless telecommunication networks designed specifically for IoT applications. They allow for long-range communication at a low bit-rate among connected items, such as battery-powered sensors. However, with these benefits come also a number of drawbacks, including the limited data rate available and the reliance on low power channel access methods which can negatively impact performance in a highly dense network.

The purpose of the research contained in this work is to measure the performance in terms of Quality-of-Service (QoS), Packet Delivery Ratio (PDR) and scalability of one LPWAN in particular, Long Range Wide Area Network (LoRaWAN), as well as providing possible improvements that current and future network owners can put into practice. LoRaWAN simple channel access protocol, based on pure Additive Links On-line Hawaii Area (ALOHA) is intended to reduce cost, complexity, and energy consumption while increasing transmission range. However, it also severely limits the scalability of the technology, making it more prone to packet collision, despite LoRaWAN being particularly resilient to self-interference, thanks to the underlining, proprietary Long Range (LoRa) modulation. In this thesis, LoRaWAN technology is evaluated through both software simulation and experimental deployments, with the goal of gaining a deeper understanding of the technology to then create better models and better performing deployments. The innovations and novel results presented throughout will accelerate the pervasiveness of LPWAN networks such as LoRaWAN, and ultimately their effectivness.

Despite being developed in 2015, LoRa and LoRaWAN have both not been fully characterised, particularly in regard to large-scale behaviour. This is partly due to the low feasibility of deploying vast networks. To address this, the first recorded instance of an urban digital twin of a 20 devices LoRaWAN network was deployed and analysed.

The available simulation models, despite being successfully used in various research studies, are also not fully complete, and a deeper understanding of the technology is

required to fix some remaining open issues.

To give additional insight into their operation as well as practical improvements that can be carried out to maximise performance, both from a consumer and an industrial standpoint, existing LoRa and LoRaWAN modules for the network simulator NS-3 are enhanced and used throughout the work presented. Scalability and Quality-of-Service improvements are also presented, based on the knowledge gaps found in current LoRaWAN research and the results of the simulations performed. In particular, improvements on PDR up to 10% are reported using novel techniques of downlink independent optimisation, and new insight on the positioning of gateways to achieve maximum scalability in a two-gateways network are also highlighted.

List of Publications

Journal and Magazine Papers

- **B. Citoni**, F. Fioranelli, M. A. Imran, and Q. H. Abbasi, “Internet of Things and LoRaWAN-Enabled Future Smart Farming,” IEEE Internet of Things Magazine, vol. 2, no. 4, pp. 14–19, dec 2019.

- **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, “Impact of Inter-Gateway Distance on LoRaWAN Performance,” Electronics, vol. 10, no. 18, 2021.

- **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran and S. Hussain, "Comparative Analysis of an Urban LoRaWAN Deployment: Real World Versus Simulation," in IEEE Sensors Journal, vol. 22, no. 17, pp. 17216-17223, 1 Sept.1, 2022.

Conference Proceedings

- **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, “Comparative analysis of discrete time simulations and stochastic geometry models of a single gateway lorawan network,” in 2021 IEEE International Conference on Smart Internet of Things (SmartIoT), 2021, pp. 8–12.

- **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran and S. Hussain, "Downlink Independent Throughput optimisation in LoRaWAN," 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), 2022, pp. 1-5.

List of Acronyms

3GPP 3rd Generation Partnership Project.

ADR Adaptive Data Rate.

AI Artificial Intelligence.

ALOHA Additive Links On-line Hawaii Area.

BER Bit Error Rate.

BW Bandwidth.

CF Centre Frequency.

CO_2 Carbon Dioxide.

CR Coding Rate.

CRC Cyclic Redundancy Check.

CSS Chirp Spread Spectrum.

DC Duty Cycle.

DFT Discrete Fourier Transform.

DPSK Differential Phase-Shift Keying.

DR Data Rate.

ED End-Device.

EIRP Effective Isotropic Radiated Power.

EIS Electrical Impedance Spectroscopy.

EKG Electrocardiogram.

EPP Energy Per Packet.

EV Electric Vehicle.

EWL External Wall Loss.

FEC Forward Error Correction.

FFT Fast Fourier Transform.

GPS Global Positioning System.

GUI Graphical User Interface.

HTTP Hypertext Transfer Protocol.

ICS-LoRa Interleaved Chirp Spreading LoRa.

IEEE Institute of Electrical and Electronics Engineers.

IH Implicit Header.

IIoT Industrial Internet of Things.

IoT Internet of Things.

ISM Industrial, Scientifical, Medical.

ITS Intelligent Transportation System.

LoRa Long Range.

LoRaWAN Long Range Wide Area Network.

LoS Line-of-Sight.

LPWAN Low Power Wide Area Network.

LTE Long-Term Evolution.

MAC Medium Access Control.

MIC Message Integrity Code.

MQTT Message Queue Telemetry Transport.

NB-IoT Narrowband IoT.

NF Noise Figure.

NS Network Server.

NS-3 Network Simulator 3.

OFDM Orthogonal Frequency-Division Multiplexing.

PDR Packet Delivery Ratio.

PHY Physical.

PL Path Loss.

PLR Packet Loss Rate.

PM Particular Matter.

PPP Poisson Point Process.

PSK Phase-Shift Keying.

QoS Quality-of-Service.

QPSK Quadrature Phase-Shift Keying.

RSSI Received Signal Strength Indicator.

RTLS-UWB Real-Time Localisation System based on Ultra-Wide Band technology.

S-ALOHA Slotted-ALOHA.

SC-FDMA Single-Carrier Frequency-Division Multiple Access.

SEM Simulation Execution Manager.

SER Symbol Error Ratio.

SF Spreading Factor.

SINR Signal to Interference & Noise Ratio.

SNR Signal-to-Noise Ratio.

SSK Slope-Shift Keying.

SSK-ICS LoRa Slope-Shift Keying, Interleaved Chirp Spreading LoRa.

TDMA Time Division Multiple Access.

ToA Time-on-Air.

TSCH Time Slotted Channel Hopping.

UTI Urinary Tract Infection.

V2G Vehicle-to-Grid.

V2I Vehicle-to-Infrastructure.

V2V Vehicle-to-Vehicle.

WSN Wireless Sensor Network.

List of Symbols

G_{rx} Receiver Antenna Gain.

G_{tx} Transmitting Antenna Gain.

G_t Total Antenna Gain.

$L^{dB}_{buildings}$ Building Penetration Loss.

$L^{dB}_{propagation}$ Propagation Path Loss.

PL^{dB}_0 Path Loss at Reference Distance d_0.

PL^{dB} Path Loss.

PL_b Payload Bytes.

PL_{sym} Payload Symbols.

P_r Received Power.

P_t Transmitted Power.

$Preamble_{sym}$ Preamble Symbols.

R_{chips} Chip Rate.

R_{sym} Symbol rate.

S_g Gateway Sensitivity.

S_n Node Sensitivity.

S Sensitivity.

T_{air} Packet Time-on-Air.

$T_{payload}$ Payload Time-on-Air.

$T_{preamble}$ Preamble Time-on-Air.

$T_{silence}$ Time of Silence.

T_{sym} Symbol time.

ToA_{MaxSF} Time-On-Air of packets using the maximum allowed SF.

$Tx_{interval}$ Transmission Interval.

X_g Fading Factor.

γ Path Loss Exponent.

ω Standard Deviation Factor.

ψ Wavelength.

ρ Correlated Shadowing.

c Speed of Light.

d_0 Reference Distance.

dd_0 Decorrelation Distance.

d Link Distance.

f_c Carrier Frequency.

f frequency.

h_{re} Effective Receiver Antenna Height.

h_{te} Effective Transmitter Antenna Height.

h Antenna Height.

z Relative Gateway Distance.

Contents

List of Tables

List of Figures

Acknowledgements

I am grateful to my supervisors Dr. Sajjad Hussain and Dr. Qammer Abbasi for the support and the help throughout my PhD and the UK EPSRC supporting this PhD studentship financially.

I am also forever thankful to my former supervisor Francesco Fioranelli for taking me on all those years ago and Prof. Muhammad Imran for being an ever-present figure for all of us in the CSI group.

Speaking of which, I need to extend a big thank you to all the colleagues who made these years fly by. Aman, Abed, João, Jarez, Paulo for inviting me in their exclusive lunch group and making me part of the gang. Shuja, Yusuf, Kayode for the knowledge they tried to pass on to me. Jingke and Haobo for introducing me to the glamorous world of hotpots and "see you tomorrow" mushrooms. Federico for the unsolicited Italianness. Giuseppe and Maria for helping to wrap my head around stats and probabilities. The countless football people whom I managed to play with over the course of these years. Paesano for all the pizza. My friends, colleagues and life-long companions from Rome: Lorenzo, Francesco, Gabriele, Iacopo, Andrea, Silvia, Pier Paolo, Carlo Alberto, Adriano, Valerio for half a lifetime of friendship.

Finally, I thank my family. My mum, who so badly wanted me to have a third and hopefully last graduation ceremony, my dad for paving the way to become a Dr. and my sisters for their love and support.

Declaration

With the exception of Chapters 1, 2, 3 and Sections 4.1 which largely contain introductory and background material, all work in this thesis was carried out by the author unless otherwise explicitly stated and referenced.

Chapter 1

Introduction

1.1 Research Background

Since its widespread adoption over the past decades, the Internet has revolutionised the world. Today its presence permeates multiple aspects of everyday life, a result of which is the fact that we live in a constantly connected world. With improving technology and decreasing costs, everyday objects have also gained the ability to connect and exchange information with the Internet. Such devices form what is collectively referred to as the "Internet of Things" (IoT), widely regarded as the next paradigm shift in computing [1], after the advent of the World Wide Web.

The main benefit of Internet-equipped objects is that they allow for intelligent decision-making based on the system status they are monitoring, coupled with some user-defined AI. For this reason, IoT applications to specific areas are also colloquially called "smart" e.g. smart city [2] and smart agriculture [3]. With the increasing number of connected devices and specific sensors being developed, IoT technology is gaining traction in a plethora of both consumer applications and multiple industries, such as healthcare [4], automotive [5, 6], oil and gas [7] and manufacturing [8], as seen in Fig. 1.1.

It is forecast that the total amount of connected devices will increase to 6.2 billion by 2030 [9]. Aside from the financial possibilities stemming from the adoption of such a rapidly expanding technology, the introduction of smart decision-making can help to solve problems and limitations of current production in ways that are beneficial to society. For instance, in the agricultural industry, IoT technology can be leveraged to bridge the gap between current food production and the estimated production needed in the future. World population is expected to grow by 2.3 billion people before 2050 with most of this growth predicted in developing countries. As a result, food production in these countries will need to almost double [10] and increasing production alone may not be enough to achieve food security. Applications such as smart irrigation systems can utilise water more efficiently, optimising electricity consumption and cost of labour; sensors on plants and soil

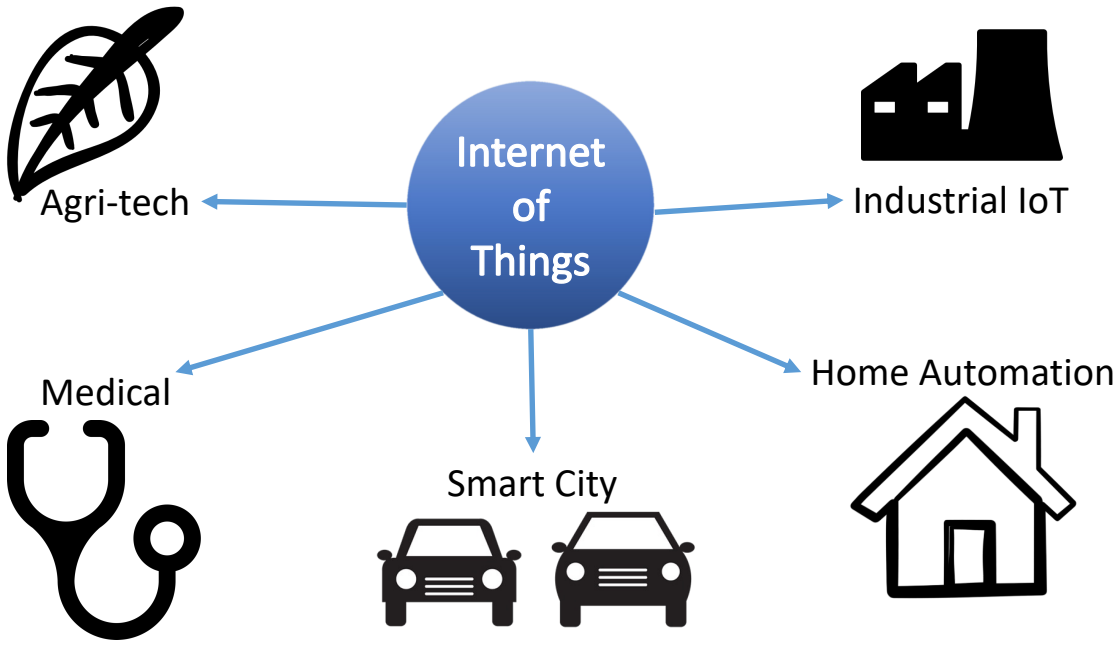


Figure 1.1: Internet of Things verticals

can optimise the delivery of nutrients and increase yields. The same technology can also be applied to numerous other industries to grant similar improvements in performance.

Because of their nature, IoT networks need to be based on a communication technology that satisfies different requirements:

Long Achievable Range: With devices placed over vast areas that can span over kilometres, such as agricultural fields or industrial warehouses, IoT networks require a communication protocol that can reliably connect over vast distances [11], possibly without requiring repeaters or multi-hop functionality.

Low Power Consumption: Sensing devices in an IoT network are generally assumed to be battery-powered. Because of the varied applications, devices in a network could be needed to be deployed in hard-to-access areas. Due of this, they should be setup to be "deployed and forgotten", meaning they should be capable of operating for as long as possible on a single battery charge with little to no maintenance [12].

High Scalability: Whether covering kilometres or a single building, it is conceivable that a single receiver could service a high number of transmitters, especially in urban deployments, which tend to be denser. Given this possibility, IoT networks need a protocol that is highly scalable, i.e. one where the network quality does not dramatically decrease as the total number of connected devices increases.

Low Cost: Given the previous considerations with regard to the number of possible devices deployed, a typical LPWAN device should be as inexpensive as possible, to allow for the development of profitable applications. This is both done at a capital cost level, by reducing a device complexity and features, and at an operating cost level by using

unlicensed, free to use, frequency bands such as the Industrial, Scientifical, Medical (ISM) bands in the EU, which becomes an implicit spectral requirement alongside the maximum radiated power.

These requirements are not fully met by any of the traditional communication protocols, be they short-range, like Bluetooth or Wi-Fi, or long-range, like cellular and satellite. Cellular technologies are flawed by design, as they are meant to handle the high data rates of multimedia traffic by allowing only relatively few devices to connect to each base station while granting them a wide bandwidth. This is the opposite of what is required in an IoT network, where a high number of nodes only need the bandwidth necessary to transmit a few bytes every few minutes. This makes long range technologies very difficult to be scaled up without increasing costs and energy consumption. Satellite coverage, while having the best range of all the technologies, is simply too expensive and energy inefficient for multiple sensor applications [13].

Traditional short-range communication protocols such as Wi-Fi and Bluetooth suffer partly from the same design flaw of cellular. Although these technologies are in use in some IoT applications today, they were also designed to handle a higher volume of data than what is required for standard IoT purposes, at the expense of increased power consumption, which is unfeasible for battery powered devices to be used in rural and difficult to access environments.

These shortcomings were mitigated with the introduction of lower energy protocols based on IEEE 802.15.4 and designed for WSN such as ZigBee. However, their mesh, multi-hop network topologies, increasing the number of devices a signal has to travel through to reach the network server, also exponentially increase the network complexity and its power consumption with the number of connected devices. Multi-hopping topologies such as these will trade off low power consumption for a higher transmission range [14].

For large-scale, sustainable applications over vast areas, both the long-range communication available to cellular and satellite, the lower deployment costs associated with Bluetooth and Wi-Fi and the power consumption of protocols designed with wireless sensor networks in mind such as ZigBee are needed.

1.2 Research Motivation

LPWANs (Low Power Wide Area Networks) are protocols designed to complement the existing ones and to better fit the IoT ecosystem specifically. LoRaWAN, Narrowband IoT (NB-IoT) and Sigfox are all examples of LPWANs used in IoT and IIoT (Industrial Internet of Things) applications. Given the relative youth of the LPWAN technologies, it is understandable that a single one has not yet emerged as the definite best for each possible application and IoT use cases. However, one that has received particular attention from

both the academic and industry world is LoRaWAN, a solution developed and maintained by the LoRa Alliance, which was established in 2015. LoRaWAN is built on top of the LoRa proprietary modulation technique.

The gathered attention is partly due to LoRaWAN open-source nature, the promise of low capital and maintenance costs and potential long battery life with very high communication range. Moreover, LoRaWAN is very flexible, allowing for both private and public networks, and affording the designer's control over its performance via multiple trade-offs, such as increasing the communication range at the expense of battery life. This makes it suitable for a variety of specific use cases, since multiple parameters can be fine-tuned.

Given the attention and recognition LoRaWAN has been gathering it is important to assess its scalability, or how the protocol performance degrades as additional devices are connected. The manufacturers claim that a single LoRaWAN receiver can sustain potentially up to thousands of devices concurrently [15], while independent research shows this figure is very optimistic, with actual numbers realistically closer to a hundred [16].

This scalability can be studied by leveraging the different built-in ways in which LoRaWAN allows the network designer to alter its performance. In particular, the quasi-orthogonality of the LoRa modulation Spreading Factors (SFs) and the Adaptive Data Rate (ADR) algorithms are both ways that a network can trade off some of its benefits, such as low power consumption and high range for improved scalability.

However, because of the practical unfeasibility of setting up thousands of devices to carry out empirical assessments over these properties, most of the scalability analysis on the protocol is dealt with via theoretical means [17–19] and discrete-time packet level simulations [16, 20, 21]. Generally, due to the mathematical complexity of modelling a thoroughly accurate network, most theoretical research limits the scope of the analysis to simpler, less realistic configurations, usually by reducing the number of variables involved in a realistic network design. This simplification also occurs to a certain extent for discrete-time software simulations.

1.2.1 Research Question

The research contributions outlined in this thesis are based on theoretical studies conducted through software simulations as well as empirical results gathered from a physical deployment. These studies are all towards addressing the following main research question:

How can the performance of realistic and feasible LoRaWAN deployments be improved based on what we can learn about the large-scale behaviour we can assess in software simulations?

1.3 Research Objectives

Understanding LoRaWAN behaviour:
The foundation upon which the research in this thesis was built on is an improved knowledge of LoRa and LoRaWAN behaviour. The underlying LoRa modulation, which was established as a proprietary technology in 2015, has by now been successfully reverse-engineered by multiple researchers [22–24]. However, its potential, especially when applied to Medium Access Control (MAC) Layers like LoRaWAN has yet to be completely understood in the many and diverse IoT settings it could be applied to. This research served as the inspiration for a number of improvements to existing NS-3 simulation models, enabling better simulation of LoRaWAN networks.

Improving and validating simulation models:
In order to assess the large-scale behaviour of this LPWAN protocol, thorough and extensive discrete-time software simulations of LoRaWAN networks with varying key transmission parameters such as the packet's spreading factor were carried out. These were made possible following a number of modifications made to already existing simulation models of LoRa and LoRaWAN behaviour. These enhancements are reported in full in Section 4.2 and include changes to the interference matrices used to determine packet collision, the expansion of the existing channel line-up and the correction of erroneous behaviour such as a slightly shorter Time-on-Air (ToA) than expected and the inability of demodulation paths to receive packets on multiple frequencies. Validation for these modification was sought by performing a novel comparison with results from already well-established analytical analysis.

Propose novel scalability enhancements:
To study how performance could be improved in a realistic and feasible network, a way to optimise Quality-of-Service (QoS) independently of the network's feedback was also sought. For this, the novel concept of context-agnostic, downlink-independent optimisation was established and put forward in Section 5.2. Similarly, the impact of each of the packet outage conditions that can affect a LoRaWAN packet was analysed in a 2-gateway network as a function of the distance between the two gateways, in Section 5.1.

Creating the first LoRaWAN Digital Twin:
As a final objective, the first recorded instance of an accurately deployed LoRaWAN digital twin was designed to be deployed and studied. This would involve 20 end-devices to be deployed across the urban context of the University Campus, to then by modelled as accurately as possible in software simulations.

Throughout completing all these objective and analysing the results of software simulations, as well as empirical deployments, information useful to improve the capacity, scalability and QoS of LoRaWAN networks of various sizes and purposes are found and presented to the scientific community, in an effort to make IoT technology more pervasive.

1.4 Research Publications

1.4.1 Journal and Magazine publications

In the early stages of this PhD, much effort was put into deploying the research group's first LoRaWAN network. Alongside the familiarisation with the then available equipment, the initial focus was on working within the realm of smart agriculture. The first published paper reflects this:

- [3] **B. Citoni**, F. Fioranelli, M. A. Imran, and Q. H. Abbasi, "Internet of Things and
  LoRaWAN-Enabled Future Smart Farming," IEEE Internet of Things Magazine, vol. 2, no. 4, pp. 14–19, dec 2019.

This is a review paper cataloguing the current test beds and known agri-tech deployments, as well as highlighting some of the research questions surrounding LoRaWAN networks at the time of writing.

In the following year, the focus of the thesis was shifted to application-agnostic protocol research based on software simulations. This was mostly due to the onset of COVID-19, which made it extremely difficult to work on real-life deployments for the better part of two years as access to facilities, labs and, in fact, any outdoor location was severely limited.

The first published work reflecting this is:

- [25] **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Impact of Inter-Gateway Distance on LoRaWAN Performance," Electronics, vol. 10, no. 18, 2021.

This journal paper highlights how the performance of a two gateway network varies as the two gateways are moved relatively to each other, from them being hypothetically placed in the same location to the opposite ends of a circular network. This simulation-based work is the first that was carried out by leveraging the discrete time simulator NS-3 and existing LoRaWAN modelling tools, aptly modified for the occasion.

- [26] **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran and S. Hussain, "Comparative Analysis of an Urban LoRaWAN Deployment: Real World Versus Simulation," in IEEE Sensors Journal, vol. 22, no. 17, pp. 17216-17223, 1 Sept.1, 2022.

Finally, as the COVID-19 restrictions began to be lifted towards the end of 2021, work began on deploying a 20 nodes network to be set up around the University's main campus. This was used both to validate the theoretical research carried out over the previous years in simulation and also provide a comparison between the empirical network performance and the performance of an equivalent simulated network. This paper has been published on the IEEE Sensors Journal.

1.4.2 Conference Proceedings

The first conference paper that was produced dealt with establishing a comparison between the mathematical models used for LoRaWAN analysis and the output of discrete-time simulations. This was deemed particularly useful, as the two methods had never before been used in tandem but always exclusively of each other. Moreover, as NS-3 was chosen as the simulating method going forward from this point, this work also provided validation of the updated modules being used. Ultimately, both modelling methods, simulations and stochastic geometry, were found to be roughly equivalent when recreating the network under inspection, albeit with different strengths and weaknesses.

- [27] **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Comparative analysis of discrete time simulations and stochastic geometry models of a single gateway lorawan network," in 2021 IEEE International Conference on Smart Internet of Things (SmartIoT), 2021, pp. 8–12.

Another conference paper, presented in June 2022, investigates instead possible ways to improve a network throughput, scalability and QoS without resorting to downlink, whether containing control or simply information arriving from the Network Server (NS). This is particularly interesting to further the main research question of this thesis, as this represents some concrete improvement that can be easily achieved in a realistic, personal network that a user could set up, as opposed to downlink-heavy methods usually proposed in literature which are more effective but exponentially more complex.

- [28] **B. Citoni**, S. Ansari, Q. H. Abbasi, M. A. Imran and S. Hussain, "Downlink Independent Throughput optimisation in LoRaWAN," 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), 2022, pp. 1-5.

This work was presented at the 1st IEEE Workshop on Sustainable and Intelligent Green Internet of Things for 6G and Beyond at IEEE-VTC SPRING 2022, and won the Best Paper Award for the workshop.

1.5 Thesis Outline

The remainder of this thesis is structured as follows, with a breakdown of the research output linked to each chapter reported in Table 1.1.

Chapter 2 briefly presents LPWANs before focusing on the LoRa modulation and the LoRaWAN protocol.

In Chapter 3 existing literature is analysed to provide more context to the work to follow and to outline the current state-of-the-art of application specific testbeds and deployments. Research addressing the protocol performance and scalability is also analysed to point out current gaps in knowledge.

Table 1.1: Published papers relevant to each chapter

Chapter	Related Publications
1	Internet of Things and LoRaWAN-Enabled Future Smart Farming [3]
2	-
3	-
4	Comparative analysis of discrete time simulations and stochastic geometry models of a single gateway lorawan network [27]
5	Impact of Inter-Gateway Distance on LoRaWAN Performance [25] & Downlink Independent Throughput optimisation in LoRaWAN [28]
6	Comparative Analysis of an Urban LoRaWAN Deployment: Real World Versus Simulation [26]
7	-

Chapter 4 introduces the network simulator NS-3 and how LoRaWAN networks have been effectively modelled and simulated through its use. The original source and a brief explanation of the simulation models is given before a list of the needed modifications that we carried out to better suit the research carried out and presented in this thesis. Finally, the soundness of the analysis performed via NS-3 is validated against results from mathematical modelling of LoRaWAN obtained through stochastic geometry, as published in [27].

Chapter 5 dives into the novel research concerning the scalability of LoRaWAN, discussing the simulation set up and outcomes of published research [25, 28].

Chapter 6 presents the deployment of a 20 end-devices network that was set up throughout the University's grounds, its real-life performance as well a comparison with the simulated performance of the digital twin that was developed in NS-3.

Finally, Chapter 7 includes concluding remarks, an overview of the contributions to the field, and the suggested next steps for future work around the subject.

Chapter 2

Technical Background

This chapter includes notions about LPWANs and specifically LoRa and LoRaWAN that will be helpful and referenced through the rest of this thesis.

2.1 Low Power Wide Area Networks

A Low Power Wide Area Network (LPWAN) is a type of wireless wide-area network that allows long-distance communication, often in the range of kilometres, while also having a very low power consumption [29]. Despite sharing a single goal, the various LPWANs that have emerged on the market over the last decade present some key differences that make them better suited to specific applications and environments.

A first distinction can be made between ultra-narrow band and wideband LPWANs, based on the range of frequencies the signal is modulated on. This modulation can also be performed in different ways, including Chirp Spread Spectrum (CSS) [30], Differential Phase-Shift Keying (DPSK) [31] and Quadrature Phase-Shift Keying (QPSK) [32] modulation. Despite having key differences such as CSS being a spread spectrum modulation and DPSK and QPSK being Phase-Shift Keying (PSK), digital modulations they all rely on using low modulation rates to encode more information and energy into each bit, increasing the coverage. [29]. In PSK, a finite number of phase shifts are used to encode the digital signal onto the carrier while in CSS wideband and linear frequency modulated sinusoidal chirps are used to encode information [33]. The technology might be either open-source or proprietary and, additionally, operating on licensed or unlicensed bands of the frequency spectrum. Licensed protocols transmit on the licensed frequency spectrum and as such have a higher cost, but fewer restrictions imposed on their transmissions.

Cellular-based LPWANs usually operate on these frequencies. Their main advantages are that their duty cycle is not restricted by spectrum-sharing regulations and that the needed infrastructure required to give country-wide coverage is already in place, as cellular communication is ubiquitous. However, this benefit is often offset by the higher operating

cost and higher power consumption associated with wideband, high data-rates cellular communication.

One example of Cellular-based LPWAN is NB-IoT, operating on licensed frequency bands and developed by the 3rd Generation Partnership Project (3GPP). It is a derivation of the Long-Term Evolution (LTE) standard, but with specific changes that make it more suitable for IoT purposes and in particular high-density applications, with each cell capable of supporting around 50000 devices. Among these are the restriction of the LTE bandwidth from 1.4-20 MHz to 180 kHz and lower power consumption achieved at the expenses of data-rate and latency. NB-IoT is a half-duplex protocol, using Orthogonal Frequency-Division Multiplexing (OFDM) for downlink transmission and Single-Carrier Frequency-Division Multiple Access (SC-FDMA) for uplink [34, 35].

On the other hand, unlicensed protocols make use of the unlicensed portions of the spectrum, such as the sub-GHz ISM frequency bands. These are free to use frequencies, but strict regulations are in place so that the spectrum can be fairly shared among all the different devices and applications using it. Most notably, duty cycle limitations may be put into place, limiting the amount of total transmitting time each device is allowed in a certain time span.

Sigfox is a proprietary, ultra-narrowband technology, operating on the unlicensed frequency spectrum. Specifically, the 868MHz band in Europe and the 915MHz band in the USA, part of the sub-GHz ISM bands. The ultra-narrowband approach provides extensive coverage at the expense of the data rate. Sigfox restricts the amount of uplink to 140 per device per day, and the payload size for these to 12 bytes. Similarly, downlink is restricted to 4 messaged with 8 bytes of payload. Because of Sigfox being owned by the homonymous company, which is the only network provider, this protocol has a larger coverage and it is more available worldwide than other proprietary LPWANs [14, 35].

2.2 LoRa

LoRa is a proprietary Physical (PHY) layer modulation based on Chirp Spread Spectrum (CSS) modulation, originally developed by Cycleo and acquired in 2012 by Semtech [36]. Given its proprietary nature, the exact specifications and functioning of the LoRa modulation are not officially documented. Despite this, the full PHY chain was successfully reverse-engineered over the past years by different researchers, based on a combination of public patents and experiments [22–24].

2.2.1 Chirp Spread Spectrum-based Modulation

Chirp Spread Spectrum (CSS) modulation has found extensive applications in radar systems throughout the years [37]. CSS is based on the premise that a chirp signal, which

is a sinusoidal signal with a linearly variable frequency and constant duration, may be used to "spread" information over a larger spectrum than it would otherwise need. At the cost of lower spectral efficiency, this uniform distribution of a symbol over a larger bandwidth provides resistance to frequency-selective noise and interferers. This makes this modulation more resistant to multi-path interference and the Doppler effect. This, in turn, increases the transmission range by allowing receivers to be less sensitive, but at the cost of a lower data rate.

With respect to standard CSS, LoRa modulation modifies the chirps by encoding multiple data bits in each chirp, as well as applying a starting frequency shift which is related to the number of bits encoded. The number of bits, also called chips, that are encoded in each chirp is represented by the Spreading Factor (SF) of the transmission. Specifically, a chirp using SF N will encode 2^N chips per symbol, and there will be 2^N possible starting frequencies for the chirp. A chirp is designed in such a way to start from a frequency f, contained within the available spectrum. It then ramps up in frequency until the end of the available band and wraps around to f from the band's lower end. Multiplying this up-chirp with its complex conjugate will yield a pure sinusoid, which frequency can be obtained through a Discrete Fourier Transform (DFT) and which will represent the original, frequency shifted, data.

2.2.2 PHY Packet Structure

As far as the PHY layer is concerned, an uplink packet always starts with a programmable preamble, used for synchronisation with the receiver. A LoRa preamble is made up of a minimum of eight base frequency up-chirps. The amount of the first up-chirps may be programmed and is used to detect the start of each frame. The programmable section is followed by two up-chirp symbols that encode the sync word needed for frame synchronisation, network identification, and preventing gateways from accepting packets from another LoRa network. Finally, the preamble concludes with two down-chirp symbols utilised for frequency synchronisation. After the last two symbols, a 0.25 symbol time signifies a silent time intended to allow the recipient to align the time [38–40].

The PHY header can be implicit or explicit, though LoRaWAN specifications mandate it to be the latter for every transmission [41]. In this case, the header contains the payload length in bytes, the Forward Error Correction (FEC) code rate of the payload and the header Cyclic Redundancy Check (CRC) [42]. The header is always protected with the FEC of the highest code rate of 4/8 [38, 43].

The payload can contain LoRaWAN MAC layer control packets as well as data, and will be further analysed in the next section. It can be followed by the payload CRC, though only for uplink transmissions.

A visual representation of an uplink packet is in Fig. 2.1.

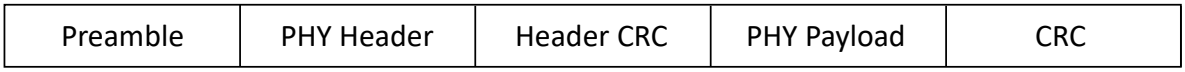


Figure 2.1: Structure of an uplink LoRa PHY packet

2.2.3 Transmission Parameters

As mentioned in the introduction, LoRa has received particular attention as the modulation technology to enable IoT deployments due to its customisable parameters, allowing it to suit different applications with ease. The following parameters are extracted from the available documentation on Semtech's SX1272 and SX1276 transceiver chips [44, 45].

The **Spreading Factor (SF)**, as mentioned in the previous section, relates to the number of data chips that are encoded in each frequency chirp. Using a SF of N will result in up-chirp that is 2^N chips long. Incrementing the SF cuts transmission rate in half, doubling transmission duration and, ultimately, energy consumption. However, increasing the duration of the chirps also improves the Signal-to-Noise Ratio (SNR) of the transmission by extending the time it takes a packet to be received [46], making the choice of spreading factor a trade-off between coverage, data rate and power consumption, as shown in Table 2.1. The spreading factor of a LoRa transmission can be set between 6 and 12. Using SF6 requires, among others, packets having implicit PHY headers, hence it is not used in LoRaWAN.

The **Bandwidth (BW)** is the frequency range in the transmission band over which the chirp spreads. Higher BW results in a higher data rate (and consequently shorter transmission time), but a reduced sensitivity due to the incorporation of extra noise. Using a lower BW grants a better sensitivity but a lower data rate. The data is transmitted at a chip rate equal to the bandwidth. As a result, a 125 kHz bandwidth equates to a chip rate of 125 kc/s. The most commonly used bandwidths are 125kHz, 250kHz and 500kHz (BW125, BW250 and BW500, respectively).

The **Coding Rate (CR)** refers to a programmable number of bits that are added to the packet header to perform FEC. CR can be set between 1 and 4 (CR1 to CR4), resulting in an FEC rate of $4/(\text{CR} + 4)$. Larger coding rates increase resilience to interference, but also increase packets length, air-time and energy consumption [47].

The **Centre Frequency (CF)** defines the central frequency of transmission. This is typically in the sub-1 GHz band for LoRa.

The **Transmission Power** which is the radio power used by a transmission. LoRa chipsets support a maximum range of powers between -4 and 20 dBm, in 1dB steps.

2.2.4 Derived Parameters

The main features, parameters and potential issues with the LoRa modulation can be derived and discussed using the above transmission parameters as a starting point. These are also contained in the documentation for the SX1272 and SX1276 chips [46, 48–50].

Symbol Duration

The overall time duration of a single LoRa chirp, or symbol, is given by the following:

$$T_{sym} = \frac{2^{SF}}{BW} \ seconds \tag{2.1}$$

and thus is dependent on both the SF and the BW used for transmission. The symbol rate can then be calculated simply as the reciprocal of T_{sym}:

$$R_{sym} = \frac{1}{T_{sym}} = \frac{BW}{2^{SF}} \ symbols/seconds \tag{2.2}$$

And the chip rate is:

$$R_{chips} = R_{sym} \ \times \ 2^{SF} \ chips/seconds \tag{2.3}$$

Packet Length and ToA

The time-on-air of a LoRaWAN packet can be calculated via the following [50]:

$$T_{air} = T_{preamble} \ + \ T_{payload} \ seconds \tag{2.4}$$

with $T_{preamble}$ and $T_{payload}$ being the time duration for the preamble and the PHY payload, respectively.

Knowing a symbol's duration from 2.1, we can calculate $T_{preamble}$ and $T_{payload}$ as the product of their length in symbols with T_{sym}.

$$T_{preamble} = (Preamble_{sym} + 4.25) \ \times \ T_{sym} \ seconds \tag{2.5}$$

with $Preamble_{sym}$ being the number of programmed preamble symbols, 8 for the UK [51], and T_{sym} being a symbol's duration.

$$T_{payload} = PL_{sym} \ \times \ T_{sym} \ seconds \tag{2.6}$$

with PL_{sym} being the number of symbols for the PHY payload.

$$PL_{sym} = 8 + max(\lceil \frac{8PL_b - 4SF + 28 + 16CRC - 20IH}{4(SF - 2DE)} \rceil \times (CR + 4), 0) \ symbols \tag{2.7}$$

Table 2.1: SNR and Gateway Sensitivity for different combinations of SF and BW [44, 46]

SF	BW	SNR (dB)	S_g (dB)
7	125 kHz	-7.5	-124.5
8	125 kHz	-10	-127
9	125 kHz	-12.5	-129.5
10	125 kHz	-15	-132
11	125 kHz	-17.5	-134.5
12	125 kHz	-20	-137

with:
PL_b = number of frame payload bytes (1 to 255), to be added to the 13 additional mandatory LoRaWAN overhead bytes;
SF = Spreading Factor;
IH = 0 if header is explicit, like it is mandatory for LoRaWAN, 1 if implicit;
CRC = 1 if optional 16-bits to perform CRC are present, 0 otherwise;
DE = 0 or 1, depending on whether low data rate optimisation is present. This depends on the T_{sym}. If greater than 16 ms then DE must be 1;
CR = Coding Rate from 1 to 4 (being 4/4+CR).

Receiver Sensitivity

The sensitivity of a radio receiver can be calculated with the following:

$$S = -174 + 10log_{10}(BW) + NF + SNR\ dB \tag{2.8}$$

where:
S is the receiver sensitivity;
-174 is caused by thermal noise and can only be changed by altering the receiver's temperature;
NF is the Noise Figure for the receiver and is typically taken as 6 dB for LoRa transceivers [46], though it depends on the receiver architecture,
SNR is the Signal-to-Noise Ratio required by the underlying modulation scheme, and it is dependent on the SF of the transmission, as shown in Table 2.1 [44, 46].

SF Orthogonality

One of the main advantages of the LoRa modulation is that the different spreading factors allow the receiver demodulator to discriminate between incoming concurrent transmis-

sions on the same frequency channel, making the different spreading factors theoretically orthogonal to each other. This circumvents some of the limitations imposed by using a pure ALOHA channel access MAC such as LoRaWAN, namely the high chance of packet collision and the resulting loss of all packets involved typical of pure ALOHA networks. In reality, it has been proven that the different spreading factors are only quasi-orthogonal to each other, meaning that each transmission will also suffer interference from packets sent at the same time, on the same frequency, but on a different SF, albeit to a lesser extent [39, 52, 53]. This is expanded upon when discussing the different outage conditions for a LoRaWAN packet, in Section 2.4.

2.3 LoRaWAN

LoRaWAN is the de facto MAC layer protocol built upon LoRa and its specifications are openly available. It was developed and is currently maintained by the LoRa Alliance, a non-profit organisation comprised of more than 500 telecommunications corporations, universities, and manufacturers. LoRaWAN's full technology stack is visible in Fig. 2.2.

Other custom protocols can be built upon the LoRa PHY layer, partially trying to address or mitigate some of LoRaWAN limitations. LoRaBlink, was developed by [54] to achieve multi-hop, robust, and low-latency communication, while keeping a low energy profile. Symphony Link is another protocol that aims to resolve the problem of scalability outlined by different researchers by implementing a range of different features [55].

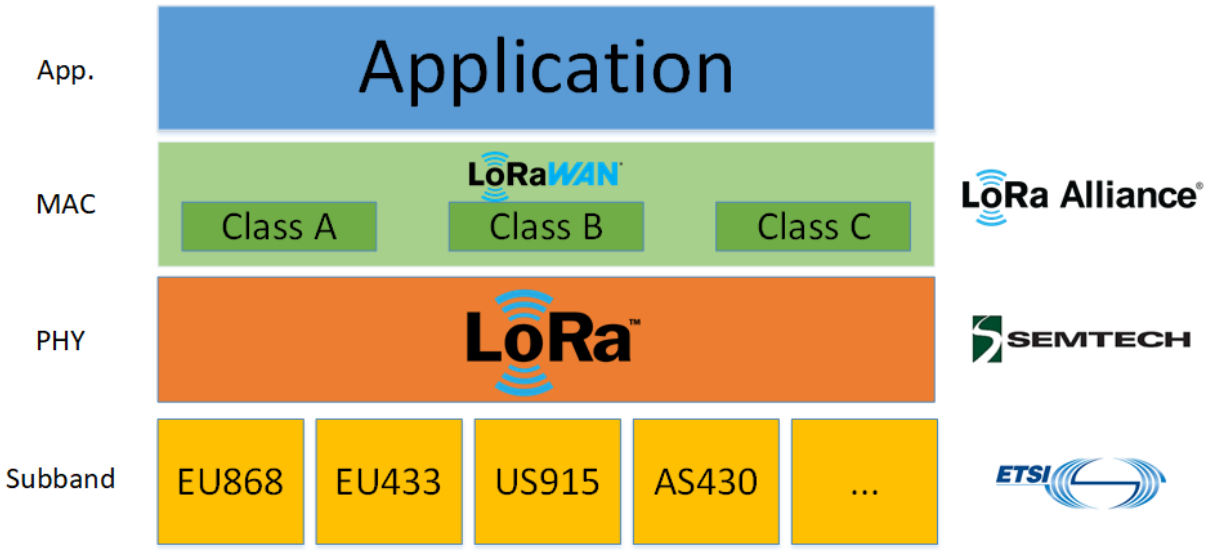


Figure 2.2: LoRaWAN technology stack, from [56]

2.3.1 Network Topology

LoRaWAN networks are comprised of three main elements:

- Nodes or End-Devices: electronic devices, typically including sensors, responsible for collecting data or implementing instructions via actuators through communication with one or more gateways.
- Gateways: Internet connected receivers that forward the packets coming from the nodes to a NS, acting as a logically invisible bridge between nodes and network.
- The NS, which handles the de-duplication of received packets, rejection of corrupted/unwanted ones, as well as scheduling messages to be sent to specific nodes through gateways in range.

A typical LoRaWAN network, as it is common for LPWANs, is organised in a star-of-stars topology akin to the one shown in Fig. 2.3, with nodes transmitting to all gateways in range rather than having a direct link to any single gateway.

The star-of-stars topology has its main advantage in requiring a single hop to achieve uplink and downlink communication, potentially saving power. Multi-hopping topologies, typically utilised by a Wireless Sensor Network (WSN), effectively trade off power efficiency in order to achieve higher range using multiple clusters of devices communicating to a sink they have to associate with directly, and which in turn forwards the message on.

Thanks to this limitation, the estimated lifetime of a single battery-powered LoRaWAN connected sensor device is expected to be years, which results in cheaper deployment and maintenance as well as an overall simplified network design [12, 13]. The maximum recorded range achieved by an uplink message using LoRaWAN is 832 km [11].

The gateway then relays the data it receives from all the nodes in range to the network server associated with each node. Finally, the NS forwards them to another element present in most networks, the Application Server, a server operated by the end-user to processes the data. As this is user-specific and optional, it is not included in the network topology in Fig. 2.3. The management of security keys for all the linked devices is handled by a separate Join Server [57].

Communication is bidirectional, so devices can send data to the network via uplink, and receive instructions via downlink. Uplink direction is however strongly favoured. Communication between two nodes is not available with LoRaWAN, requiring the data to pass through a gateway in both uplink and downlink for this to be achieved. Downlink communications must be sparse, since gateways are subject to the same duty cycle limits as nodes. Additionally, downlink also affects the overall QoS of a network by adding additional interference and preventing gateways from receiving uplink messages while busy

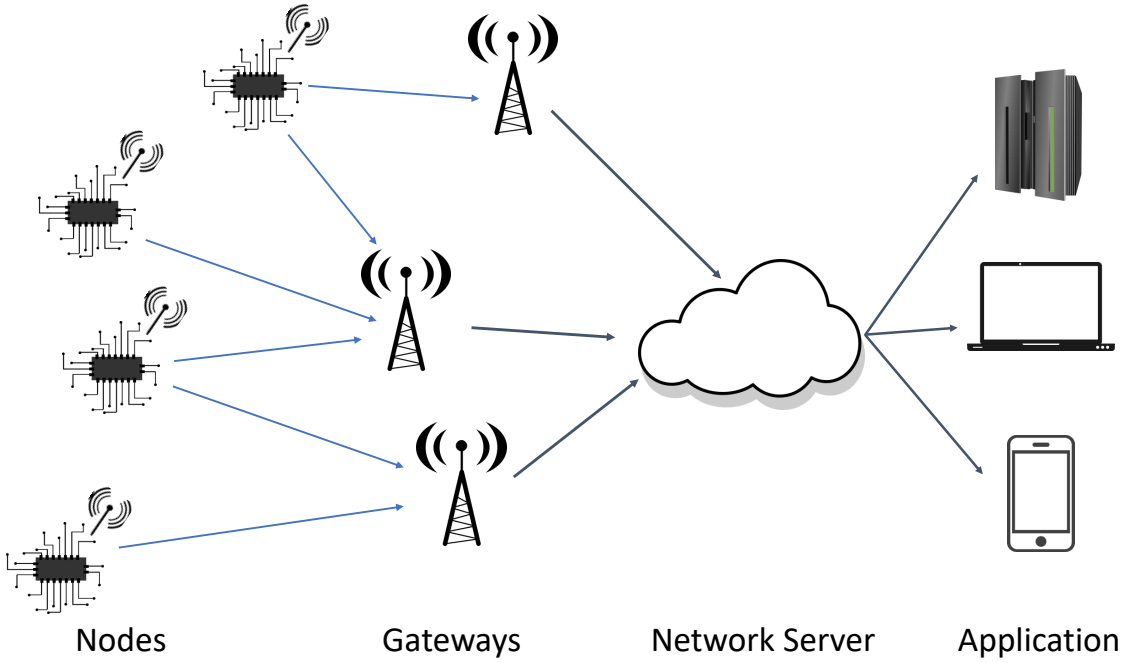


Figure 2.3: Typical star-of-stars LoRaWAN network topology

with downlink frames. These limits prevent the use of LoRaWAN for time-critical, low-latency applications.

LoRaWAN defines its data rate as a function of the spreading factor and bandwidth used by the underlying LoRa modulation for a specific transmission. A decrease in SF or an increase in BW always represents an increment on the Bit Rate achieved, as shown in Table 2.2.

2.3.2 Packet Structure

The content of the PHY Payload of an uplink packet, which structure was broke down in Fig. 2.1 above, is itself reported in Fig. 2.4, below.

The MAC Header specifies the message type and format of the transmission. There are 8 possible message types, defined by a 3 bits number: Join-request, Rejoin-request, Join-

Table 2.2: LoRaWAN Data Rate (DR) per combination of SF and BW in EU863-870 [51]

Data Rate	SF	BW (kHz)	Indicative Bit Rate (bit/s)
0	12	125	250
1	11	125	440
2	10	125	980
3	9	125	1760
4	8	125	3125
5	7	125	5470
6	7	250	11000

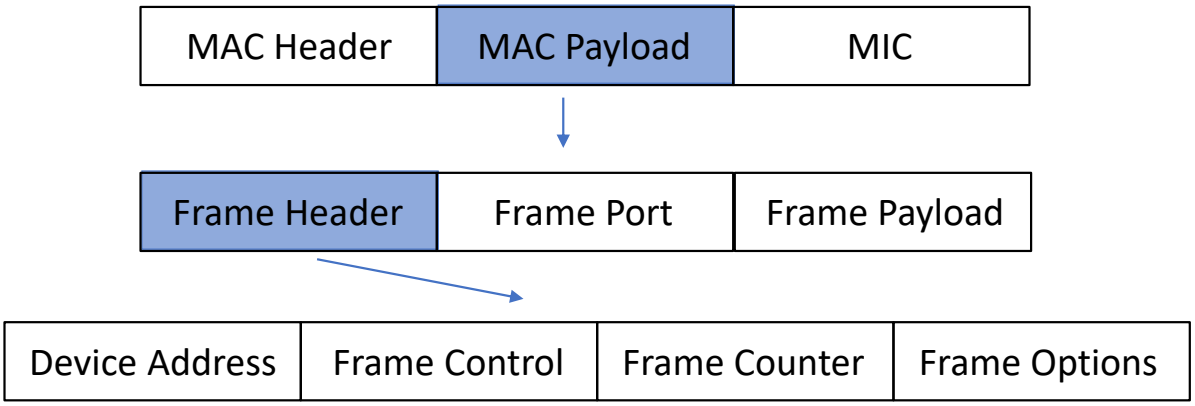


Figure 2.4: Content of PHY Payload from Fig. 2.1

accept, unconfirmed and confirmed data up and down and proprietary protocol messages [41].

Other important information is also contained within the 13 bytes LoRaWAN mandatory overhead such as the device address, a frame counter and information on whether the message requested acknowledgement or usage of the Adaptive Data Rate (ADR) algorithm.

2.3.3 Device Classes

LoRaWAN networks define three different classes for its end nodes, Class A, Class B and Class C [57, 58].

- Class A is mandatory to implement for every LoRaWAN device. At any time, a Class A device can send an uplink message, initiating communication. Following the completion of the uplink and a short delay, the devices open two mutually exclusive, short time windows where messages can be received in downlink. These two windows (RX1 and RX2, respectively) have different settings to maximise the chance of either of them picking up the preamble of the downlink message. Class A supports bidirectional communication, but downlink can only be received after an uplink is sent. This allows Class A devices to be very energy efficient, as, aside from while transmitting and the two short receive windows, they can be in deep sleep [57].

- Class B devices are provided with the same downlink capabilities of Class A, plus extra slots for downlink communication at periodic times, called “ping slots”. These ping times are set through downlink communication via periodic, time-synchronising beacons every 128 seconds. Because they are reachable at predetermined intervals and can receive downlinks without completing an uplink first, Class B end devices have lower latency than Class A nodes. However, they also have a higher power consumption [57].

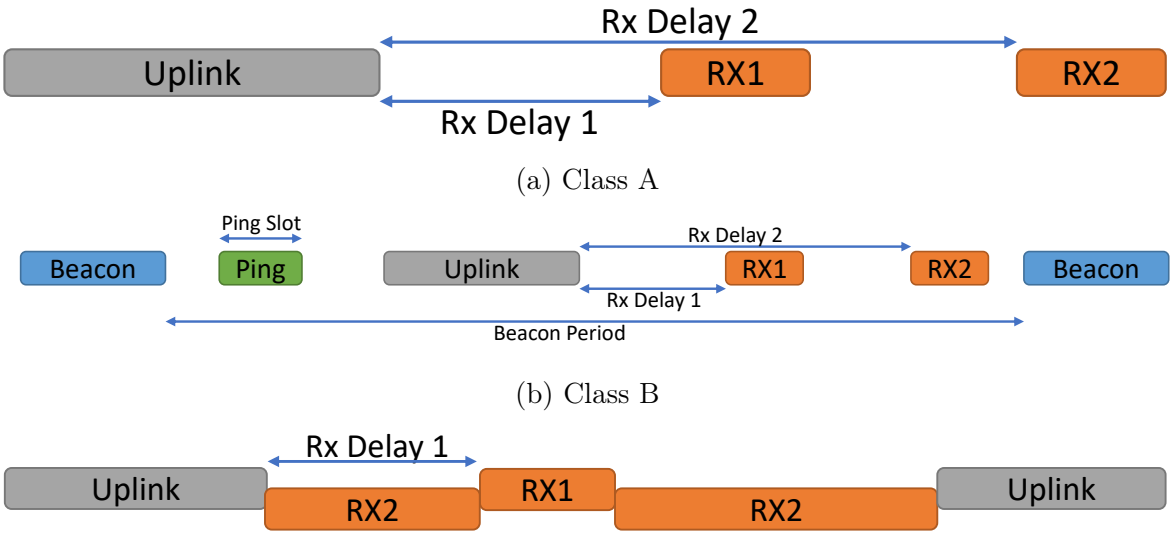
- Class C is meant for mains-powered devices, with little to no energy constraints. A Class C device continuously listens for downlink messages, except when transmitting [47]. Low-latency communication is possible, but at a prohibitive energy cost for battery operation [57].

Fig. 2.5 shows the timing diagrams for the three classes of devices. While research has been extended to other classes, particularly Class B [59,60], Class A remains the most popular and used of the three. Most of the available LoRaWAN research, as is this work, is focused on Class A devices.

2.3.4 EU Regulations

LoRaWAN devices are expected to be operational throughout the world. The main benefit of using the unlicensed spectrum is the lower operating costs resulting from the shared use of the ISM bands. Using lower frequencies also theoretically provides deeper wave penetration than higher frequencies in the range of GHz. However, ISM bands, such as the most commonly used EU 863-870 or the US 902-928 MHz also have continent and even country-specific regulations. These mandate specifications for both the LoRaWAN MAC and the underlining LoRa PHY layer and can create problems for LoRaWAN devices, which strive to be as consistent as possible around the globe.

LoRaWAN operation is usually defined across a number of frequency bands and sub-bands. In EU, in addition to the standard 863-870 MHz band, LoRa can also be set to operate in the lower ISM band of 433 MHz [51] or in the higher 2.4GHz ISM band [61].



(a) Class A

(b) Class B

(c) Class C

Figure 2.5: Timing diagrams for the different possible classes of LoRaWAN devices

The EU863-870 MHz band is further subdivided in sub-bands h1.4 (865-868 MHz) and h1.5 (868-868.6 MHz) [62, 63].

Restrictions required by the ISM bands include definitions for the packet preamble, the mandatory channel frequencies, the maximum payload size, the reception windows, the allowable spreading factors, the maximum radiated power, the maximum duty cycle per device and the Join processes. The following restrictions apply to the 863-870 MHz frequency band commonly used in the EU, though similar restrictions apply regardless of the operating range, with perhaps slightly different parameters.

Two of the most important and limiting factors when it comes to the scalability of LoRaWAN technology are the limit in maximum radiated power and the duty cycle restriction.

Each transmission on the EU 863-870 MHz frequencies needs to be within the approved Effective Isotropic Radiated Power (EIRP). This is the total radiated power by an isotropic antenna in a single direction. In the UK the maximum power allowed is 25 mW, equivalent to 14 dBm. Both sub-bands in the 863-870 MHz range also have a strict duty cycle of 1% [64, 65].

Each sub-band must remain "silent" for a period of time that is proportional to the ToA of the packet and the maximum available duty cycle enforced [66]. For instance, for an air-time of 1 second and a duty cycle of 1%, each sub-band will have a 99 seconds mandatory silence time. This restricts the available air-time per device to roughly 36 seconds per day, which makes the technology unsuitable for high data rate applications. LoRaWAN devices usually can transmit over multiple channels and utilise channel-hopping algorithms to choose, pseudo-randomly, the best possible channel to transmit the data on in order to try to mitigate this drawback. The only restriction is that channel utilisation must not cause the sub-band's duty cycle to go above the mandated limit. Having 2 sub-bands effectively doubles the amount of possible air time from 36 seconds to 72 [67].

Aside from three mandatory frequencies on which transmission must be enabled on all LoRaWAN devices, the network operator has complete control over channel allocation in the EU 868 MHz band. The mandatory frequencies are 868.1, 868.3 and 868.5 MHz. The transmission channel is chosen pseudo-randomly, with only restriction that channel utilisation must not cause the sub-band's duty cycle to go above the mandated limit, which is usually 1%. Most readily available commercial gateways feature 8 possible channels, which are programmable by the user. However, these are usually set to coincide with those available from the network provided. An example is the channel line-up for The Things Network [57], which is included in Table 2.3.

Table 2.3: EU Frequency Plan for [57], including relevant sub-bands [62, 63], maximum EIRP and duty cycle limit [64, 65]

Frequency (MHz)	Sub-Band (MHz)	Max EIRP (dBm)	Duty Cycle
868.1 868.3 868.5	868-868.6	14	1%
867.1 867.3 867.5 867.7 867.9	865-868	14	1%

2.4 LoRaWAN Outage Conditions

Among the various metrics that can be used to analyse the performance of a wireless network, of paramount importance is the amount of data that is correctly received among all messages sent. In this thesis this is expressed as the Packet Delivery Ratio (PDR) of a network, which is calculated, as a percentage, with the following:

$$PDR = 100 \times \frac{packets\ received}{packets\ sent}\ \% \tag{2.9}$$

LoRaWAN applications span many use-cases and can give rise to networks of vastly different sizes. Although the ideal 100% PDR is an unrealistic goal in the majority of real-life scenarios, a unanimously ”acceptable" PDR is not defined in literature, depending, rather than on the protocol, on the application itself.

A packet in a LoRaWAN network can be lost for a number of reasons, which henceforth will be referred to as Outage Conditions. The following conditions, partially retrieved from this thesis's author work on [25], need to be true during the lifetime of a packet to achieve correct transmission. A single failure along the chain will cause data to be lost either because of the packet being dropped, not properly decoded or not being sent in the first place. Therefore, they should always be taken into account when analysing and modelling LoRaWAN behaviour.

2.4.1 Duty Cycle Limitations

As mentioned, the enforced duty cycle on the commonly used frequency sub-bands of 863.00–868.00 MHz and 868.00–868.60 MHz for LoRaWAN in EU is 1%. Each sub-band has to observe a period of silence between transmissions that is proportional to the time-on-air of the last transmitted packet and the maximum available duty cycle enforced,

according to the following [66]:

$$T_{silence} = T_{air}(\frac{1}{DC} - 1)\ seconds \tag{2.10}$$

where:
$T_{silence}$ = time-of-silence required after transmission;
T_{air} = time-on-air;
DC = Enforced Duty Cycle.

Since the transmission interval between packets ($Tx_{interval}$) is set by the designer, possible transmissions happening over the duty-cycle limits are usually not considered as an outage condition. However, should the parameters of transmission vary for any reason, such as a ADR algorithm being enabled, a device might try to transmit over its duty cycle limit. As an example, a device set to transmit a packet of 10 bytes every 90 seconds on a single sub-band works as expected using DR5 (SF7 and BW125) with an air-time of 61 ms and $T_{silence}$ of 6.1 seconds (from Equations (2.10) and (2.4)). If the same device was to switch to DR0 (SF12 and BW125) without adjusting its $Tx_{interval}$ it would now have a ToA of 1482.8 ms and $T_{silence}$ of 146.8 seconds, above 1% duty cycle. LoRa transceivers are usually set up to abide to this limit, so the transmission would be stopped by the chip itself [68]. This results in "invisible" packet loss, as this loss would not be visible by inspection of the node's frame-counter alone. The data will not be queued up for retransmission unless the device's behaviour has been specially designed to do so, and the data stored in the packet would be lost. This behaviour can be manually overwritten, but doing so may violate the spectrum access laws.

2.4.2 Transmission Below the Receiver's Sensitivity

All wireless signals are affected by a certain amount of power loss during travel through any medium, including large-scale loss, or Path Loss (PL), and small-scale loss, or fading. The former characterises the average signal intensity that reaches a receiver and is generally distance-based, increasing as the transmitter and receiver are placed further apart. Not only is the signal weakened by propagation, but reflections, scattering, and diffraction all contribute significantly to the average Received Signal Strength Indicator (RSSI). This is known as fading, and it results from the superposition of different copies of the signal with various phases and amplitudes. This is usually described as a random process, because of the interaction's inherent complexity [33].

Nodes and gateways can only receive and correctly decode packets whose receive power is above a certain power threshold, also known as a device sensitivity, shown in Table 2.1. If a packet fails to maintain a certain strength, it will be lost.

2.4.3 Packet Collision and Interference

LoRaWAN is based on a pure ALOHA channel access scheme. In its simplest form, this allows devices with data to send to do so at any time without checking for availability, scheduling conflicts with other devices or even a successful connection. Because of this, colliding frames resulting from simultaneous transmissions will be damaged and likely lost. A network operating using a pure ALOHA has a maximum theoretical throughput of 18% [69, 70].

This theoretical maximum is increased by the usage of LoRa and its patented modulation scheme due to the quasi-orthogonality of the spreading factors and the capture effect.

Each LoRaWAN packet will experience interference from both external sources and other devices operating in the same network. Assuming perfect orthogonality between the different spreading factors used in the LoRa modulation, only packets sent at the same time, on the same frequency and using the same SF will cause interference to one another. However, the SFs' quasi-orthogonality has been demonstrated to have a non-negligible influence on overall network quality and scalability, with each packet also facing interference from packets transmitted at the same time, on the same frequency, but on a different SF [39, 52, 53].

The effect of this is somewhat lessened by a strong capture effect which has also been found to apply to LoRa transmissions. In the event of a collision the stronger signal received has a good probability of still being successfully demodulated if it achieves a minimal power difference from the strongest interferer. These thresholds have been historically modelled according to the work performed in 2015 by [71]. The authors show how a packet will be received over interference on the same SF if it is at least 6 dB stronger. More recent developments in research conducted by [39] show different thresholds which are obtained with an experimental setup. In this case, only 1 dB separation is required for the strongest signal to be correctly received over the total interference on the same SF. These are currently regarded as the standard for LoRa. The rejection matrices from both studies are reported in Table 2.4.

Moreover, time overlap between two packets is not always destructive. Experiments in [43] have shown that during packet collisions caused by concurrent transmissions, one of the packets will be received if the final six symbols of the preamble and header do not overlap [72].

2.4.4 Demodulator Channel Saturation

LoRa packet reception entails a receiver detecting the 4 consecutive upchirps in a packet's preamble [73] and "latching on" to it, provided the receive power is above the receiver's

Table 2.4: Inter-SF and Intra-SF destructive interference levels in dB in [71] and [39]

(a) Thresholds from [71]

	SF7	SF8	SF9	SF10	SF11	SF12
SF7	-6	16	18	19	19	20
SF8	24	-6	20	22	22	22
SF9	27	27	-6	23	25	25
SF10	30	30	30	-6	26	28
SF11	33	33	33	33	-6	29
SF12	36	36	36	36	36	-6

(b) Thresholds from [39]

	SF7	SF8	SF9	SF10	SF11	SF12
SF7	-1	8	9	9	9	9
SF8	11	-1	11	12	13	13
SF9	15	13	-1	13	14	15
SF10	19	18	17	-1	17	18
SF11	22	22	21	20	-1	20
SF12	25	25	25	24	13	-1

sensitivity. On a typical LoRa gateway, if reception is initiated, the operation will take up one of the receiver's demodulating paths for a time proportional to the packet's ToA, after a time delay based on the distance between sender and receiver. Commercially available LoRa gateways can usually receive up to 8 packets simultaneously, thanks to 8 parallel demodulators [74]. If all demodulating paths are occupied by packets being currently decoded, the next packet coming across such a gateway will not be "latched on", and so it will not be received.

2.4.5 Simultaneous Communication

Standard operation for LoRaWAN devices is half-duplex. Nodes and gateways can communicate with each other, but not at the same time, meaning no packet reception is possible while transmission is occurring. This also implies an implicit priority to be established between uplink and downlink. This would be used to decide if the reception of an uplink packet should stop in case a gateway requires to transmit a downlink message or not. No formal rule is presented by the LoRaWAN standard, but it is accepted that the most popular and widely researched gateway chipset, the SX1301 [74] will stop incoming packet reception in favour of transmitting [75, 76].

2.5 Adaptive Data Rate (ADR) Algorithm

The ADR utilised by LoRaWAN is a method for optimising the data rate, energy consumption and total air-time of packets in a network, with the overarching goal of improving network scalability and its QoS. Choosing the best data rate for transmissions in fact improves the number of devices that can be handled by a gateway. Usage of the ADR is not mandatory, but is recommended by the LoRaWAN standards [41].

The ADR is responsible for updating the spreading factor, bandwidth and transmission power of transmissions initiated on any one device and is divided into two distinct algorithms, one of which operates on each node device where it is enabled, and the other on the NS.

On the node side, the algorithm is formally defined in the specifications [41]. If enabled, the algorithm will decrease the data rate of the node following lack of acknowledgement of packet reception, thus extending the range. This process is assuming that by doing so, a better link margin can be established between the node and a gateway, which possibly prevented packet reception and the requested acknowledgement. This side of the ADR is only responsible for decreasing the data rate, making transmissions slower, less energy efficient but more resilient and increasing their range.

For the algorithm operating on the network server, which is instead responsible for increasing the data rate, there is no official formulation. However, a baseline method is documented in [77], which has been reportedly adopted by open-source network operators such as The Things Network, and has been the basis for discussion for research working on the development of new and improved ADR algorithms, as highlighted in Section 3.2.5. A known limit of the baseline ADR that these developments try to overcome is for instance its suitability for static devices only, for which the radio channel does not vary consistently with each transmission.

When a node transmits an uplink message, it may be received by a number of different gateways. The message is then tagged with its Received Signal Strength Indicator (RSSI) and SNR at each gateway, before being deduplicated and sent to the application server. The message metadata will then include separate information about the "goodness" of the link between the transmitting node and each gateway the packet was received by.

After collecting a number of uplink messages, 20 according to the suggested method, the network server uses the collected information to determine the most suitable data rate for each device by comparing previously received Signal to Interference & Noise Ratio (SINR) values to the current data rate (including SF and BW) and transmission power level.

The ADR is then responsible for updating the spreading factor, bandwidth and transmission power of all subsequent uplinks by transmitting the new ones to be used via a downlink message to the device.

2.6 LoRaWAN Large-scale Propagation Loss Models

There are several ways to define the Path Loss (PL) in a specific environment, but this is always represented as the ratio of the received power P_r to the transmitted power P_t. In this section the models used for describing losses in a typical LoRaWAN environment, including those used in this thesis authors' publications will be discussed.

The free space model is the most basic for describing distance-based propagation loss. It assumes that there is a direct Line-of-Sight (LoS) between devices and that the signal propagates in a straight line. The PL in the free space model is expressed as [78]:

$$PL^{dB} = \left(\frac{\sqrt{G_t}c}{4\pi d f_c}\right)^2 \; dB \tag{2.11}$$

where:
G_t is the total gain of the antennas;
c is the speed of light;
d is the link distance;
f_c is the carrier frequency.

This purely analytical approach to PL definition is not often used to describe LoRaWAN propagation, as LoS is not easily achieved other than for satellite-based and some microwave links.

Instead, the majority of radio propagation models which are commonly used are developed by a combination of analytical and empirical methods. The empirical methods have the advantage of implicitly accounting for all propagation parameters, known and unknown, via real field observations. However, they can only be applied to that specific environment, or similar ones, before new measurements and new models have to be constructed [33].

These models are the most frequently used when dealing with urban and suburban environments. One such model is that developed by Okumura in 1968 [79]. It applies to frequencies ranging from 150 MHz to 3000 MHz and distances ranging from 1 km to 100 km. It is suitable for base station antenna heights ranging from 30 to 1000 metres. The Hata model [80] is an empirical representation of Okumura's path loss, and it is valid from 150 MHz to 1500 MHz. The formula provided can be altered, and different formulations can be obtained to model urban environments of different sizes, as well as suburban and rural areas. For urban regions, the formula for median path loss is [33]:

$$PL^{dB} = 69.55 + 26.16 log_{10}(f_c) - 13.82 log_{10}(h_{te}) - a(h_{re}) + (44.9 - 6.55 log_{10}(h_{te})) log_{10}(d) \; dB \tag{2.12}$$

where:
f_c (MHz) is the frequency of the transmission (between 150-1500 MHz);
h_{te} (m) is the effective height of the transmitting antenna (between 30 and 200 m);
h_{re} (m) is the effective height of the receiving antenna (between 1 and 10 m);
d (km) is the link distance;
$a(h_{re})$ is a correction factor which is a function of the size of the coverage area.

Finally, another popular formulation is the Log-Distance model. This can be used to model both indoor and outdoor scenarios and follows from the realisation that whether indoor or outdoor, the average received signal power drops logarithmically with distance.

$$PL^{dB} = PL_0^{dB} + 10\gamma log_{10}(d/d_0) + X_g \; dB \tag{2.13}$$

with:
PL_0^{dB} (dB) is the path loss at the reference distance d_0
d is the distance between receiver and transmitter;
d_0 is the reference distance;
γ is the path loss exponent;
X_g is the fading factor, a zero-mean Gaussian distributed random variable (in dB) with standard deviation ω (also in dB), modelling log-normal shadowing.

This model is very flexible as d_0, γ and X_g can statistically describe the path loss for any arbitrary location, given a certain link distance d. As such is a popular choice for network simulators. Additionally, the values for γ and X_g can also be calculated from empirical data using linear regression [33].

Chapter 3

LoRaWAN Applications and State of The Art

Based on the flexibility of the protocol as highlighted in the previous chapter, LoRa and LoRaWAN have the potential to be applied effectively to a number of different use cases. The impact of the different SFs especially, with the trade-off between power efficiency and coverage arising by varying the SF of a transmission, is one of the main enabler of LoRaWAN for different use-cases. These include smart agriculture, smart cities, industrial, smart homes, smart healthcare and smart metering [81]. In this chapter we will, firstly, in Section 3.1, be giving an overview of the different applications of the technology over the various IoT verticals, as highlighted in Fig. 1.1. This is done by analysing test-beds and proof of concept studies which showcase the breadth of possible application of the technology and are aimed at justifying the research interest and effort that LoRa and LoRaWAN have gathered since their introduction in IoT. In the following Section will then put into focus the application agnostic research which is most relevant to the work presented in rest of this thesis, especially the academic research concerned with improving the scalability of LoRa and LoRaWAN.

3.1 LoRaWAN Applications in IoT and IIoT

In this section the most common IoT and IIoT application for a range of use-cases are presented. While not directly connected to the protocol's scalability, these academic papers are important to showcase some of the common pitfalls of practical LoRaWAN usage, such as the limited availability of a handful of researchers to develop large-scale networks, the lack of perfect knowledge of each device featured in a network and the range of current propagation loss models. This information was crucial to formulate the research questions presented in this thesis's introduction, and fed into the protocol knowledge used to work on the enhancements to the simulation models which are presented in the upcoming chapters

of this thesis.

3.1.1 Smart Agriculture

In the agri-tech, or smart agriculture context, sensor nodes are typically concerned with the welfare of livestock and conditions of plants and soil. Uplink messages make up the bulk of a network's data, monitoring factors that only need updating every few minutes. This effectively works around one of the main drawbacks of the LoRaWAN technology, namely its maximum data rate limit. Downlink messages are present, though only sporadically used. In a typical scenario, they could be used to activate simple devices such as solenoid valves and switches, like a sprinkler to water a specific portion of a field or a dispenser to refill livestock food and water troughs [82].

There are multiple instances of LoRaWAN network operating in the agri-tech vertical, both commercially and at a research-level. Some of the most common applications of LoRaWAN in agri-tech which are currently being researched and developed include irrigation control, optimisation of water usage, soil monitoring and chemical analysis, pest and diseases control and prevention, machine and agricultural manned and unmanned vehicles (drone) monitoring and control, movement and habits monitoring to screen and prevent diseases like lameness in livestock [3, 83].

The work performed in [84] propose a LoRa-based smart irrigation system. This system is made up of a LoRaWAN node, a solenoid valve and a hydroelectric generator. Through their experimental testing the author's suggest that the communication distance between the irrigation node and gateway can reach 8 km and that the proposed system is more efficient and reduces the cost of deployment and maintenance of such a use-case. Similarly, in [85] the authors seek to reduce labour cost while reducing water consumption with the usage of a LoRaWAN-enabled Power Switch. The findings of this study indicate the viability of deploying the LoRaWAN-based smart agriculture applications up to 1.1 km from a gateway.

In [82] the authors describe a drip irrigation monitoring system including actuators to distribute water accordingly and a Graphical User Interface (GUI) to track the operation of the network. The cost-effective solution is mostly investigated in terms of its energy consumption which is found to be low enough to give an estimated lifetime of more than 2 years per device.

In [86] the authors consider energy-efficient IoT strategies for agri-tech use-cases in greenhouse environments. To do so, they deployed a LoRaWAN based network in a greenhouse research centre in Belgium to monitor crops by recording light, temperature, CO_2 and humidity information. Three different enclosure types were tested and energy consumption as well as data visualisation was provided by using an online dashboard. They found that it is critical to deploy the sensors in an air-circulated box for precise data

and a longer lifetime.

The work presented in [87] focuses on mushroom cultivation and production in dedicated "mushroom-houses". Several sensors are placed in these structures, measuring CO_2, temperature and humidity. These are then used to activate control devices such as humidifiers and fans. This allows for the environmental factors to stay as constant as possible and ideally increase the yield of the culture.

[88] proposes applying LoRaWAN to the monitoring of the health of trees. Detection of disease is performed by Electrical Impedance Spectroscopy (EIS) analysis and results are forwarded to a central processing system via LoRaWAN. The proposed system is able to perform EIS analysis and periodically send the generated data to the central node, according to tests performed in a laboratory constant environment.

3.1.2 Smart Cities

Another sector that has enjoyed a period of growth recently is that of smart cities. As the range of possible application is very broad, possibly including anything from smart metering to vehicle communication, the requirements for smart cities use-cases vary from the more relaxed constraint akin to those of smart agriculture, to the low latency requirements typical of smart health applications. LoRaWAN has been used in a number of ways in this context. Localisation-based applications are most popular with a different range of target objects, from monitoring the position (as well as other factors) of vehicles for public transport and traffic evaluation, to assessing seating and parking availability and otherwise indoor building occupancy. Other uses include monitoring of environmental factors such as the air quality surrounding built-up areas and smart metering of power grid usage.

In [89] the authors propose, in a simulated environment, the incorporation of drones in an Intelligent Transportation System (ITS) to be used as LoRaWAN gateways. This is supposed to enhance the existing Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) systems typically utilised in such scenarios and increase energy efficiency and network resilience. The simulated results suggest that this approach can conserve between 39.2 and 12.6% of the total network energy and can improve the area stress by 52–89.7% for different amount of drones deployed over a fixed area.

[90] proposes an indoor geolocalisation for ITS based on LoRaWAN. The authors postulate that given the unfeasibility of using the more ubiquitous satellite-based localisation methods indoor, exploiting LoRaWAN metadata can provide a way of accurately measuring the position of vehicles in a location such as an underground parking. By deploying multiple gateways and using trilateration, they show that the position of a vehicle can be obtained within a 20-30 m range.

In [91] the authors propose a clustering-based technique for vehicle localisation in urban

environments. The localisation is also based and verified using the publicly available RSSI measurements collected in large urban areas in the city of Antwerp, Belgium. The solution relies on a first layer of K-Means clustering followed by a finer localisation step. In their work they are able to achieve a median error of 158.41 m and a mean error of 346.03 m based on the RSSI data, which is reportedly the best accuracy obtained through the use of the same dataset.

LoRaWAN feasibility for localisation is also explored in [92] where the authors propose a joint Global Positioning System (GPS) and Real-Time Localisation System based on Ultra-Wide Band technology (RTLS-UWB) system to provide indoor and outdoor positioning to LoRaWAN nodes. Evaluation of the obtainable performance in a real-world scenario is performed using a test setup at the University of Brescia and confirms the suitability of the proposed approach.

[93] studies a proposed hybrid architecture based on LoRa and 3G for monitoring operation in electricity and gas distribution networks. Experimental results obtained across varying distances show that Line-of-Sight (LoS) is the most determining factor in preserving a good communication link, as the presence of vegetation and buildings lower the maximum range.

In [94] the authors explore the topic of Vehicle-to-Grid (V2G) communication, between a charging station and an Electric Vehicle (EV) in motion in Pamplona. The network is characterised by building a 3D urban model and an in-house developed algorithm estimating received power levels, SNR and time domain parameters. The network's performance was then tested with moving vehicles communicating to public EV charging stations, with a 100% packet success rate when using LoRaWAN transceivers.

Authors in [95] develop a LoRaWAN testbed for air quality monitoring system in La Plata, Argentina. Three different models of sensors as well as a low volume sampler of Particular Matter (PM) are used to determine PM_{10} and $PM_{2.5}$. The results obtained show that a similar LoRaWAN deployment can be a viable tool for the determination of contaminant levels and other metrics relevant to the study of air quality.

Similarly, [96] develops a scalable system to monitor dust, Carbon Dioxide (CO_2), temperature and humidity in Yangon, Myanmar. Although only comprised of a single LoRaWAN node and a single gateway, the system also features a machine learning model trained on the air quality parameters collected. This allows to make predictions and take preventive action to alleviate the impact of pollution, and it is intended as a pioneering LoRaWAN deployment upon which more can be built in future projects in the region.

3.1.3 Remote Health

While not as popular as the previous two categories, a number of LoRaWAN based healthcare applications and papers have also been surfacing from the early years of the technol-

ogy [97, 98]. The reduced suitability of the protocol for applications of this kind is to be sought in the time-critical nature of some of the possible use-cases. It is fair to expect that a device involved in the monitoring of a person's vital signs should be the most reliable and prompt with its communication as possible. As discussed in Chapter 2, Class A LoRaWAN devices, by far the most popular available for commercial deployments and research alike, suffer from high latency due to only being able to receive downlink data following a successful uplink message. Class B to a certain extent, and Class C specifically will be more suited for these applications when they will start to become more used and widespread [57, 58]. However, Class A devices can still prove useful, especially when coupled with other technologies.

In [99], LoRaWAN is evaluated as a potential complementing technology for existing indoor and outdoor elderly positioning solutions. Outdoor localisation is performed by satellite-based methods, whereas indoor localisation is performed by Ultra Wideband technology. The resulting data is encapsulated in a LoRaWAN message and sent every minute. The network used is the same that was developed by researchers in [6] and results showed sub-meter errors in a typical indoor scenario, with an average communication latency of 700 ms.

The authors in [100] develop a point-of-care biofluid analyser for personalised monitoring using a LoRa/Bluetooth-enabled electronic reader. The study includes monitoring over different locations and an analysis of the resulting communication channel. Results prove that Urinary Tract Infection (UTI) test results are diagnosed and correctly transmitted to a remote, secure cloud server. Furthermore, the validity of an IoT-enabled medical diagnostics system was confirmed by the deployment in residential settings over distances between 1.1 and 6 km.

In [98], the authors describe a new IoT-based health monitoring system gathering medical sensor data, including blood pressure, glucose levels and temperature. Results show that with a gateway installed outdoors, at an altitude of 12 metres, the average area covered is around 33 km^2, while using at least 10 times less power than competing long-range cellular systems.

The author in [101] builds on their previous remote Electrocardiogram (EKG) Monitoring work by integrating into a LoRaWAN-based system that transmits the resulting data over long distances. At the lowest level, individual devices would be available in patient homes, where health data can be collected without the need for medical personnel. The data will then be sent via LoRaWAN to an application that can sort it and inform doctors who can further analyse the stored data.

In one of the earliest analysis on LoRa Technology, the authors [97] investigate its suitability for heath and wellness monitoring. The measurements were taken on the main campus of the University of Oulu in Finland, over an indoor area of roughly 182,400 m^2.

The measurements were carried out using a battery-powered end device strapped to a researcher's arm and reporting the sensed data to a gateway every 5 seconds. As reported by the authors, this resulted in packets only being sent every 13 seconds to respect the duty cycle limitations that were highlighted in Section 2.4. The obtained results show that the entire campus area can be covered when using the maximum transmit power of 14 dBm and the largest spreading factor of 12 on the 868 MHz ISM band, with a PDR of 96.7%.

3.1.4 Industry

Finally, the IoT paradigm, backed by LoRa technology can be successfully applied to more conventional industries such as manufacturing, construction, mining and logistics.

In [102] the authors assess the coverage area and the performance of LoRaWAN in a real indoor industrial environment, namely a flower auction warehouse. In this use-case a large number of trolleys must communicate with a gateway while moving across a large warehouse, which is made up of approximately 250000 m^2 of covered area plus additional open areas. A single gateway was installed indoors, sideways, and 6 m above the auction floor. The end-device used to perform the testing was kept at the same height as the trolleys during the measurements, about 1.7 m.

The authors in [103] focus on the practical implementation of a LoRaWAN network intended to monitor structural health in the construction domain. Two sensing node prototypes were built and wirelessly powered using a far-field wireless power transmission technique and only one dedicated RF energy source operating in the ISM 868 MHz frequency band. These sensing nodes can simultaneously measure temperature and relative humidity and wirelessly transmit the measured data over long distances and concrete using the LoRaWAN protocol.

In [104] the authors also explore LoRaWAN as the technology to provide reliable data communication in construction processes, with a deployment in Germany. The network on the construction site is remotely connected with a robotics lab and the network performance was validated through an intra-site logistics and task scheduling system. The paper demonstrates that the network behaved as desired and as such this type of use-case is open to LoRaWAN. However, they also remark that the limitations of LoRaWAN always apply, limiting suitable applications to small volume and high-latency use cases.

Finally, in a use-case embracing both industry and remote health, [105] develop a LoRaWAN-based Safety and Health Monitoring system for coal miners. The system employs smoke sensors, respiratory sensors, and heartbeat sensors to obtain environmental and health parameters. In the proposed system, a smart alert system is used to ensure the worker's safety by alerting them in the event of an accident.

3.2 LoRaWAN Research

Perhaps of more importance to the field is the different avenues of research that are available to study the behaviour of LoRaWAN as a communication technology agnostically of use-cases. Of particular interest for this thesis are those works that are attempting to formalise and improve the behaviour in large-scale deployments, tackling LoRaWAN's known scalability issues. Some of the literature tries to work directly on the modulation technique employed by LoRa by either improving the existing modulation or by proposing new, more robust, versions. Another approach is to deploy real-life LoRaWAN networks so that their performance, relative to the PHY layer parameters used, can be studied, analysed and improved empirically. However, because of the complexity and the practical issues surrounding the deployment of a network with thousands of devices, which would be required to appreciate the effects of all possible outage conditions, the pursuit of a generalised model to study such scenarios theoretically has been the most popular option. In this regard, both discrete-time simulations, performed in software, and analytical models based on stochastic geometry have been used to model the behaviour of large-scale LoRaWAN networks [38].

3.2.1 Modulation-based Research

A popular way of enhancing the standard LoRa modulation is by improving the basic, underlining CSS modulation.

In their work, authors in [106] evaluate the performance of novel modulation schemes based on CSS by experimenting with various parameters to maximise throughput and spectrum efficiency. The performance of the various CSS versions is evaluated by inspecting the Symbol Error Ratio (SER) as a function of the SNR, as well as the maximum attainable throughput.

In [107] the authors propose an extension for the standard LoRa modulation to reduce its energy consumption. This version of the modulation encodes extra data in the phase shift using Phase-Shift Keying (PSK) modulation, giving rise to PSK-LoRa. The goal is to encode more data per unit of time while maintaining the same level of performance, resulting in a more energy-efficient system. When studying the effects on the bit error rate and packet error rate, results show that in some circumstances, PSK-LoRa may transmit the same amount of information as LoRa while utilising less time resources and with no performance degradation.

Interleaved Chirp Spreading LoRa (ICS-LoRa) is proposed and described by researchers at German University in Cairo through a number of papers, including [108–110]. The goal of ICS-LoRa is to increase the number of transmitted bits per chirp signal by introducing supplementary, interleaved chirps. The suggested design preserves the communication

resilience of standard LoRa while also decreasing ToA by increasing the number of bits per symbol by one. This is proven to increase capacity of a network with only a negligible degradation in Bit Error Rate (BER). ICS-LoRa is also shown to be functioning parallel to the standard LoRa modulation, further improving on the benefits.

In their subsequent work, [111] develops a SSK modulation version, SSK-LoRa, which is shown to outperform the ICS-LoRa modulation scheme by about 1 dB SNR, and increasing the data rate compared to standard LoRa. Some of the same researchers involved would then go on developing Slope-Shift Keying, Interleaved Chirp Spreading LoRa (SSK-ICS LoRa) by integrating the two modulation modes in [112]. This mixed technique employs down chirps and interleaved up and down chirps to represent the information bits. This not only boosts the data rate of standard LoRa by up to 28.6% for the same spreading factor and bandwidth configuration, but also performs better than its competitors across a wide range of SNR per bit.

A similar proposal is also put forward by [113]. In this study, the authors suggest a novel modulation technique for LoRa communication based on combined up- and down-chirps, as opposed to the standard up-chirps of LoRa. Furthermore, a hybrid demodulation scheme based on Fast Fourier Transform (FFT) and correlation techniques is also developed to further boost performance. The performance is then assessed using simulation data and shows that the suggested approach outperforms the current LoRa modulation strategy by more than 3 dB at a BER of 10^{-4}.

Modifications to the preamble detection scheme of LoRa is also proven to be beneficial in [114]. The SX1276 LoRa transceiver is reported to need to detect 4 consecutive identical preamble upchirp signals to "latch-on" and start decoding a packet [73]. Experimental results show that relaxing this constraint from 4 to 3 non-consecutive symbols significantly improves the LoRa capture capability.

3.2.2 Empirical Performance Evaluations

As mentioned in Section 3.1, some of the research where physical LoRaWAN networks are deployed is mostly concerned with presenting purely a proof-of-concept for LoRaWAN application in that context. These include the standard IoT verticals such as smart agriculture, smart cities, healthcare and IIoT. Some however, also concern themselves with carrying a deeper analysis on the protocol itself, by performing an empirical performance evaluations of different aspects of the network. These include important metrics of network quality such as the packets SNR and the RSSI at the receiver, which are both influenced by the link quality, link distance and local interference. The SNR is a measure of the difference in power between the actual signal and the background noise, while the RSSI is the overall signal power at the receiver, both in dB. As an overarching metric, the PDR of the network, or sometimes its inverse, defined as the amount of packet lost in a network, is

most often used and reported. As opposed to standard proof-of-concept deployments, in these performance evaluation studies the emphasis is on the metadata of the transmissions rather than their data content.

These vary in scope and context but generally involve one gateway and one to few nodes [38]. Large-scale analysis is usually reserved to simulation and theoretical analysis instead.

In the previously cited work performed by [85] the author reports both information about the RSSI and SNR of the received packets in an agri-tech environment and the PDR achieved by a node in different locations against the distance from a common gateway.

In an early application of LoRa, the authors in [115] develop a trough water level monitoring system. The idea is to help cattlemen living 1-3 km away from barns to know when each trough needs to be refilled in the most straight forward way possible. During laboratory tests, the system was evaluated and deemed to be operating successfully. However, by collecting different RSSI values for packets received across different transmitter-receiver combinations it became also clear that consideration needs to go on nodes placements, particularly with regard to their elevation.

A throughout performance evaluation is carried out by [116] for the context of dairy cows monitoring. In this letter, the author characterise the path loss performance of three different barns by analysing the empirically obtained fading measurements. The results show that the large-scale fading can be described with a one-slope log-normal path loss model and the temporal fading can be statistically represented by Rician distributions with an average K-factor of 8 dB.

The authors in [117] propose and analyse two use cases for agri-tech. The first one is monitoring temperature and humidity in horse stables. The transmission of this data is set to use the minimal amount of resources the LoRa channel can use. In the second use-case, sensors for conductivity and soil temperature were buried at different depths in agricultural land, at a distance of 40 and then 350 m from a gateway. In both cases, some additional parameters are recorded, including RSSI and SNR.

[118] shows the impact of varying different communication parameters on the performance of LoRaWAN in a tree farm environment in Indiana, USA. Results show that the overall achievable range was less than theoretically predicted and that varying the PHY variables, such as the SF and the CR have a significant influence on LoRa performance. More importantly, the data dependability was uneven across distances and PHY configurations, highlighting how hard it can be to predict LoRa behaviour.

In [119], researchers from the University of Oulu install a monitoring system in a large, indoor open area of the campus spanning roughly 2163.8 m^2. The deployed system is meant to visualise the environment's real estate conditions and provides insight into understanding how a busy and crowded building can perform as a LoRaWAN IoT-enabled

building. The network includes 331 sensor nodes, each providing information about temperature, humidity, illumination, motion and CO_2. They are all programmed to have a fixed transmission period of 900 seconds, packet lengths of 24 bytes, and a spreading factor of 7. The ADR is turned off, and there is no downlink since the devices do not require confirmation of packet reception. Data received at a single gateway is gathered and analysed in terms of RSSI, SNR, and PDR, revealing that no device had less than 25% PDR while several had PDR greater than 99.5%.

[120] provides the RSSI values obtained for a LoRaWAN deployment in a three-story building in Hyderabad, India. A single node is put in different locations in the building and the resulting data is collected by a single, static gateway. This experiment aims to provide further information on how indoor impediments such as walls and doors affect LoRaWAN path loss. Similar work is carried out in [121] where the authors measure results of a deployment across multiple locations in Japan. Here too, the metadata of the packets is used to validate the path loss obtained through the empirical measurement with the theoretical one calculated on the base of the Okumura-Hata model.

In [122], the authors give an examination of a testbed placed at the University of Calgary. Two gateways are installed indoors on the 7^{th} floor of a structure, with four nodes operating at various frequencies in the 915 MHz bands. Their PDR is recorded for each floor of the building, as well as for longer exterior distances. Additional tests were conducted to determine how other characteristics, such as different spreading factors and packet sizes, impact this PDR.

The study in [123] investigates the outcomes of a large deployment in Shangai, based on data gathered over an 8-month period. The authors analyse the Packet Loss Rate (PLR) (defined as the inverse of the PDR) and attempt to identify a link between different characteristics such as the distance between transmitter and receiver, as well as the RSSI and SNR, and the PLR. The study includes two different types of nodes, each with a different transmitting behaviour. The final conclusions are made in terms of the number and density of devices in an area, the gateway deployment strategy, the different transmitting schedules, and the impact of packet collisions in the network.

Finally, in [124], the authors conduct a study across a 288 km^2 region in Brno, Czech Republic. More than 20 gateways are deployed across the city, and a single end-device is cycled through 231 test sites while operating at a spreading factor of 12 and on the required EU frequency bands. Metadata concerning the network server's timestamps, RSSI, SNR, and reached GW were retrieved and analysed in MATLAB. Throughout the trial, the total PDR was found to be 83%, and more in-depths correlations of RSSI, PDR, ADR and gateway locations are also documented.

3.2.3 Discrete-Time Simulations

Discrete-time simulation of LoRa and LoRaWAN is usually carried by software simulators, which create the ideal environment to test claims such as the 10-year expected battery life and the thousands of concurrent devices support.

There are several such simulator systems for WSNs, including NS-3, OMNET++, and SimPy. During the years, these platforms have been, thanks to the help of researches reverse-engineering the LoRa modulation, extended to enable the simulation of LoRaWAN networks. Most tools involve replicating the PHY layer behaviour, while a few also include MAC features such as ADR algorithms. [72] provides a comprehensive assessment of several LoRaWAN simulators, along with information on their native capabilities and limitations when compared to the LoRaWAN standard.

Among the most popular and widely adopted is the one developed in [16]. In this paper, the authors develop LoRaSIM [125], a custom-built discrete event simulator based on SimPy, by using insight from experiments on communication range and capture effect with varying PHY parameters. The simulator is then used to perform a scalability assessment of a network, which shows that the only way to achieve good scalability is by increasing the number of gateways and dynamically adjusting communication parameters throughout a network's life cycle. The original LoRaSIM is mostly concerned with modelling collision behaviour and as such is rather limited. This initial model was expanded further by other researchers. In [126] the authors add duty cycle restrictions and downlink capability. The LoRaWAN's ADR was added in [127] and energy consumption was added in [128]. Among other papers using this simulator are [129], [130], and [131–133].

Another popular method for performing discrete-time network simulations is to use NS-3, a network simulator written in C++.

Due to its modular nature, a number of different researchers implemented LoRa and LoRaWAN behaviour in NS-3, which has been used extensively ever since to perform all kinds of analysis regarding LoRaWAN, for example comparing its performance with that of other protocols [134].

The authors of [20] provide one of the first modules for modelling LoRaWAN. These are more comprehensive in scope than LoRaSIM, allowing for downlink traffic as well as duty cycle limitation to be observed. The authors use them to present a scalability study that includes several SF assignment techniques, varied gateway scenarios, and retransmission algorithms. They, however, seem to disregard the chance of a receiver's channel saturation.

In [135], the focus is turned to power consumption, with the module developed capable of performing power analysis alongside correct collision behaviour. However, the quasi-orthogonality of the SFs as well as the limitation of having a finite number of demodulator paths at the receiver are not considered.

A number of NS-3 LoRaWAN modules, modelling power consumption, Class B func-

tionality, ADR support and timeslot scheduling were developed and documented over a number of works [136–138] both as extension of existing models and as standalone.

One of the modules these papers extend is that developed in [21]. Here the authors develop their own NS-3 module, which is highlighted in [72] as one of the best and more complete to perform accurate PHY and MAC analysis. The module is then used to demonstrate how a gateway can ideally support up to 10^4 nodes and how increasing the number of gateways so that every gateway covers 6 km^2 can improve the coverage probability of a node to above 90%. This module has been adopted, enhanced and shared by a community of researchers ever since, including the author of this thesis.

Other simulators have also been developed and used.

In [139], the authors creates FLoRa, an open-source framework for LoRa simulations in OMNeT++. FLoRa allows for end-to-end network simulations by modelling LoRa nodes, gateways, and network servers, as well as implementing the physical and medium access control layers of LoRa, allowing for bidirectional communications and simulations of the backhaul network. The authors in [140] develop a custom-built discrete-event simulator using Java. It is intended to take into account all existing LoRaWAN capabilities and is programmed to emulate LoRaWAN operation based on North American requirements. It enables complete setup of gateways, end devices, and network settings.

In general, discrete-time simulations are easily set up and can be performed using existing libraries and software.

3.2.4 Stochastic Geometry Models

The behaviour of wireless networks and their mathematical performance have been, until recently, analysed by resorting to extreme simplifications such as only considering one or two interfering cells, or combining all interference into a single random variable that was then empirically fit to a distribution such as a lognormal [141]. Stochastic geometry is a mathematical discipline based on the modern theory of random sets by Kendall and Matheron [142] which studies models for random geometric structures and spatial patterns, based on spatial point processing. There are various point process models, most of which are based on and extend the classic homogeneous Poisson Point Process to find expressive models that allow effective statistical methods [142]. As other wireless protocols, the behaviour of LoRa and LoRaWAN has also been studied and modelled according to these stochastic geometry models.

The work done in [17] is among the first to apply stochastic geometry to simulate the behaviour of a LoRaWAN network. The authors provide a model for a single gateway network with nodes dispersed equally around it. The model assumes complete perfect orthogonality between various SFs, hence inter-SF interference is ignored. It also ignores the effects of having a finite number of demodulator paths on total outage probability and

models uplink traffic only. The same scenario, along with its constraints, is expanded by the same authors to include non-uniform node distributions in [143] and a second gateway in [144].

[19] uses a similar configuration, with a single central gateway, but this work integrates the effect of quasi-orthogonality of the SFs, using the updated collision matrix presented by [39] and reported in Table 2.4. The practical issue of having a finite number of parallel demodulation channels at the receiver, as well as downlink traffic, is not modelled. One of the few works that takes this limitation into consideration is that performed by [18]. This study examines and models a single gateway network with eight parallel demodulation paths while reducing the overall complexity of the networks in other ways, such as by assuming perfect SF orthogonality.

Most stochastic models of LoRaWAN networks limit their scope to reduced, but less realistic, configurations. This is due to the layers of complexity of a rigorously correct model and is generally accomplished by lowering the amount of variables in the network itself.

3.2.5 Scalability Research

As well as analysis and modifications to the PHY layer and the LoRa modulation, described in Section 3.2.1, the effect of changes to the MAC layer has also been investigated in literature.

Early on in LoRaWAN's research, the impact of increased gateway density and their positioning in LPWANs was defined as a key factor in the QoS of a network and its ability to scale. The authors of [145] conduct an investigation of optimal placement of 16 and 25 gateways using various tactics, and demonstrate that, while 25 gateways normally outperform 16, utilising a positioning strategy based on Fuzzy C-Means may enhance the performance of 16 gateways to equal that of 25. [146] provides both a practical and a theoretical approach for determining the optimal position of gateways in an uplink-only network. In [21] a 7.5 km radius circular urban environment was simulated in which end devices are supported by an increasing number of gateways distributed in a hexagonal grid configuration. The authors find that to achieve a reliability of transmission above 90%, gateways need to be placed in such a way that each covers 6 km^2 or, equivalently, a radius of 1200 m around it. This is similar to what was found in [16], where results indicate that the only way to scale LoRaWAN is increasing the number of gateways and dynamically adjusting communication parameters.

The dynamic parameters' adjustment is what researchers are trying to achieve by modifying and improving the existing Adaptive Data Rate (ADR) that is present on both the network server and on the end devices. Alternatively, another popular avenue is to enhance the scalability by moving away from LoRaWAN's ALOHA channel access and into

slotted-like MAC channel access. Both methods are, to different extents, downlink dependent, where either acknowledgement or parameters control is required to be transmitted to the nodes. They are also not mutually exclusive, and can be tackled simultaneously.

The authors of [147] propose FADR, a Fair Adaptive Data Rate method aimed to ease the near-far problem and provide each node with a fairer chance to be received, independent of its position relative to the gateway. Because of the capture effect, nodes suffering higher attenuation are less likely to have their packets accurately received after collision. To avoid this, FADR was developed to balance the received signal strengths from all nodes, and simulations indicate that the technique achieves greater data rate fairness than the state-of-the-art in almost all network configurations.

A similar algorithm is developed in [148], EARN, an improved greedy ADR algorithm taking into account the effect of the Coding Rate (CR), a parameter not often considered when analysing scalability. In EARN, the capture effect is leveraged to boost the survival rate of collision signals and develop the notion of an adaptive SNR margin to deal with noisy link states. In simulation, EARN surpassed standard schemes in terms of goodput, Energy Per Packet (EPP), and fairness.

[149] developed a contention-aware ADR, capable of dynamically calculating the optimal SF distribution for the network, which is then sent to each node via downlink to increase the network throughput. The number of devices that may use a certain data rate is limited, and the gradient projection approach is used to identify the best allocation. The numerical results obtained show that the proposed method is able to achieve a higher throughput than the current system.

Multiple works by a number of researchers in [131–133, 137, 139, 150–153] all present similar ideas. Changes are proposed to the standard LoRaWAN ADR that, like the original, need a certain amount of downlink to dynamically allocate LoRa settings to each node.

The channel access schemes are modifications to the ALOHA mechanism that LoRaWAN uses and is at the root of its scalability problems.

[154] proposed employing a Slotted-ALOHA (S-ALOHA) technique for LoRaWAN rather than pure ALOHA. Because the implementation is an overlay on top of the conventional LoRaWAN, no changes to pre-existing LoRaWAN firmware or libraries are needed. S-ALOHA is based on the creation of time slots for both uplink and downlink for all connected devices that are required to be time synchronised. Using the system, a decrease of 3.4x in packet collisions was appreciated in a real-world deployment of 20 nodes running for weeks.

[155] proposes a Time Slotted Channel Hopping (TSCH) approach to be applied to LoRaWAN, to better suit the technology to industrial applications. The idea is to assign a timeslot for each transmission, while also exploiting frequency and SF diversity for

scheduling simultaneous communications in those slots.

[156] performs a similar improvement by developing a new MAC layer, RS-LoRa, that aims to improve scalability and reliability by two-step lightweight scheduling implemented through downlink beacons synchronising and informing nodes. RS-LoRa is then tested in NS-3 and is shown to outperform classic LoRaWAN in terms of PDR, performance, and fairness.

The authors in [157] propose two different enhancements to the standard Class A behaviour, the Time Division Multiple Access (TDMA) and Burst schemes, to boost channel usage and system performance. The authors' calculations suggest that the proposed schemes can give more than 60% theoretical throughput, compared to the 18% maximum for the pure ALOHA scheme utilised by standard LoRaWAN. The schemes' feasibility and performance is also validated with an implementation of eight end-devices and one gateway.

In [138] the authors introduce a novel, lightweight timeslot scheduling scheme for LoRaWAN. The proposed scheme only requires a small amount of downlink transmissions and does not require time synchronisation between devices and the LoRaWAN network server. However, it is based on the fundamental assumption that devices will always transmit periodically, and as such only feasible for networks with devices behaving as such. Results from simulations carried out in NS-3 show that the scheme doubles the number of devices that a single LoRaWAN gateway can manage while keeping the same level of performance in terms of PDR.

Attempts to enhance the throughput and scalability have also been made at a network level. Charm, developed by [158] is a diversity combining scheme that allows LoRaWAN gateways to combine their received signals on the cloud, to try and decode signals that would otherwise be too weak to be correctly received and decoded by any individual gateway. Similarly, in [159] the authors developed Choir, a system that aims to exploit hardware imperfections of low-cost LoRaWAN clients to untangle and decode interfering transmissions at a single-antenna gateway. It also expands the range of each node by leveraging correlation of the sensed data. Overall, Choir enhances the throughput of the examined network by 6.84 and increases the communication range of the devices by 2.65.

3.3 Research Gaps

Based on the state-of-the-art presented in this chapter it is possible to individuate several knowledge gaps. These are going to be filled by the research put forward in the rest of this thesis.

First, the software simulation ecosystem for LoRa and LoRaWAN is currently diverse, but fundamentally incomplete. Be it LoRaSIM and its enhanced versions, modules for

NS-3, or other solutions, they are all still requiring further development to cover all the component and behaviours demonstrated by LoRa and LoRaWAN through experimental analysis.

The theoretical limits of this technology have been thoroughly researched, particularly those addressing its scalability [66]. Both stochastic geometry and discrete-time simulations are useful tools for analysing large-scale behaviour [17, 21, 132]. However, in general, these theoretical studies lack a unified behavioural model and a full list of the probable outage conditions that a packet in a network must survive in order to be appropriately received and decoded, as highlighted in Section 2.4.

Due to mathematical complexity, most stochastic geometry research restricts the scope of the study to smaller, less realistic configurations, generally by lowering the number of variables. Because both techniques are utilised in different circumstances in the literature, it is critical to prove that the results from the two distinct analytical methods are similar so that they may be used interchangeably depending on the simplicity of application in the given context.

Additionally, while analysing the scalability of LoRaWAN, the concept of performance enhancement is almost always connected with the idea of downlink feedback from the NS. Metadata including Signal-to-Noise Ratio (SNR) and Received Signal Strength Indicator (RSSI) information is processed using different algorithms and then the network informs each device of how [132, 137, 139, 150–153] and possibly when [138, 154, 156, 157] to send the next packet. While these schemes perform well in theory, the full impact on the QoS of a network due to the extensive amount of downlink required is often disregarded. Less intrusive, albeit less performing, downlink independent schemes are open to be researched.

Finally, to the best of our knowledge, no attempt to create a fully featured digital twin of a real-life deployment has been made before, instead working entirely empirically or theoretically. Despite having released in 2015, deployments and performance assessments for LoRaWAN networks are still generally constrained to small-scale, often with only one to a handful of concurrently operational end-devices, and rarely multiple gateways [38]. While large-scale networks will always be difficult to deploy, having such a digital twin would give a gauge for how effective simulations can be, allowing them to be utilised more confidently for LoRa and other LPWAN technologies.

Chapter 4

LoRaWAN Simulation Models in NS-3

NS-3 is an extension of the popular network simulator 2, and meant to model various communications networks as realistically as possible [160]. Specifically, NS-3 is a discrete-time network simulator written in C++ that can replicate various layers of the communication stack with a high level of flexibility. Additionally, the open source nature means that many technologies are already built in, with support for more to be added and expanded upon. As discussed in Section 3.2.3, a number of researchers developed NS-3 modules to replicate the behaviour of LoRa and LoRaWAN over the course of the past 5 years. The work that was performed and reported in this thesis is entirely based on the LoRaWAN module developed in 2015-16 by researchers at the Department of Information Engineering at Università degli Studi di Padova and first introduced in [161].

The following chapter includes an in-depth overview of the NS-3 simulation models introduced by [161] as well as the modifications, upgrades and changes it went through during this PhD. Finally, an initial assessment of the improved modules soundness is reported, as performed for the work presented at the 2021 SmartIoT conference [27].

4.1 The Starting LoRaWAN Module

The original model, made available on GitHub at [162] in September 2017, was built upon a number of assumptions and simplifications, necessary for making the simulation computationally viable. These simplifications included, among others, basing the system-level simulations on two models which can fully represent the transmission, the usage of a reduced channel line-up, a simplified collision model and a simplified packet receiving behaviour. The information reported in the rest of this section is included in [21] and [161], unless otherwise specified.

4.1.1 Link Measurement Model

The link measurement model is used to abstract the effects of propagation on signal strength and to average out small scale fading and other related phenomena [21]. A key element to this is the choice of a suitable channel model, to estimate the strength of a signal at a receiving side as it travels through any medium, taking into account both large and small-scale loss. The authors assume that the Path Loss (PL^{dB}) consists of contributions from both large-scale, distance-based loss, and the effect of building penetration:

$$PL^{dB} = L^{dB}_{propagation} + L^{dB}_{buildings}\ dB \tag{4.1}$$

The propagation loss, $L^{dB}_{propagation}$ used by the model is the Log-Distance one (equation (2.13)), where the values for PL^{dB}_0 and γ are computed with the following, from [163]:

$$L^{dB}_{propagation} = 40(1-4\times10^{-3}\times h)log_{10}(d|_{km}) - 18log_{10}(h|_m) + 21log_{10}(f_c|_{MHz}) + 80\ dB \tag{4.2}$$

Assuming $f_c = 868$ MHz and $h = 15$ m this becomes:

$$L^{dB}_{propagation} = 120.5 + 37.6log_{10}(d|_{km})\ dB \tag{4.3}$$

By converting the link distance d from km to m and comparing to equation (2.13), it follows that $PL^{dB}_0 = 7.7$ dB and $\gamma = 3.76$. The random effect of the fading factor X_g is not considered.

The total building penetration loss, $L^{dB}_{buildings}$ is the sum of three contributions, the external wall loss, the internal wall loss and the floor and ceiling losses and the gain in power do to height differences. These are modelled according to [164].

The External Wall Loss (EWL) for a device is modelled as a uniform random variable with a fixed range of values. This randomness approximates the range of different materials and thicknesses that exterior walls can have in various structures. The interior wall loss is taken as the maximum between two values. The first one, $Tor1$, models the loss by the internal walls by simply multiplying a loss factor, uniformly distributed between 4-10 dB and the number of internal walls in the signal path. The second one, $Tor3$, makes use of a penetration distance coefficient as well as the link distance, instead. The positive impact of elevation is modelled with the factor GFH, and increases by 1.5 dB per floor [161].

$$L^{dB}_{buildings} = EWL + max(Tor1, Tor3) - GFH\ dB \tag{4.4}$$

Finally, a model to account for correlated shadowing was also implemented, to be added to the path loss chain. This takes into account two possible correlations. Firstly, when a transmitter sends a message to a receiver, the amount of shadowing experienced

by the receiver is expected to be proportional to the amount of shadowing experienced by any other device that is "near" to it. This correlation is a function of the distance between the two devices and is often represented by an exponential function. Secondly, when two devices that are in proximity transmit, their shadowing values should correlate at the receiver [21].

The most common auto-correlation model is a decaying exponential of distance [165], and is expressed as:

$$\rho_{i,j}(d_{i,j}) = e^{-d_{i,j}/dd_0} \tag{4.5}$$

where $dd_0 > 0$ is a tuneable parameter called the decorrelation distance. For the sake of computational simplicity the module's authors chose to resort to a heuristic approach, outlined in [166]. A regular grid was built in software with each square having a side length of $dd_0 = 110$ m and an independent Gaussian random variable was drawn for each vertex of the grid. To compute the shadowing values of nodes that are not exactly situated on a grid vertex, values are interpolated using an exponential covariance function [21].

It is worth noticing that these three different models can be used together as part of a path loss chain, or they can be used singularly, especially the main, Log-Distance path loss.

4.1.2 Link Performance Model

The link performance model calculates the likelihood of receiving a packet successfully at a reduced complexity by utilising information on link strength, sensitivity and interferers via look-up tables [21].

The sensitivity for the gateways and end-devices at each spreading factor that is coded into the model is abstracted from the Semtech SX1301 [74] and Semtech SX1272 [48] chips, respectively, and is given in Table 4.1.

Any received packet whose power is below the threshold cannot be detected by a receiver. In case it is, it is assumed that the receiver will lock on the incoming signal and start receiving the packet, whose power is also assumed to stay constant for the duration

Table 4.1: Gateway and node sensitivity in the original NS-3 model [161]

SF	S_g (dB)	S_n (dB)
7	-130	-127
8	-132.5	-129.5
9	-135	-132
10	-137.5	-134.5
11	-140	-137
12	-142.5	-139.5

of the reception.

Interference for the signals is assumed to only come from other LoRa transmissions. The quasi-orthogonality of the different SFs can then be leveraged to model collision and possible surviving packets. The interference matrix reported in [71] is used, here in Table 4.2.

In the case of multiple interfering packets, the received power values for each SF are summed and checked against the packet under consideration. As mentioned before, these rejection coefficients assume that two transmissions overlap over their full length. In general, however, this is not the case, and then, the authors' of the module equalise the interfering power over the length of the transmissions that is actually overlapping, thus producing some additional leeway [21].

Finally, a single gateway is assumed to have 8 parallel receive paths, as documented in [74]. Other assumptions include: each path centre frequency must be individually configured, with multiple paths able to be set on the same frequency. Each path can receive packets on that frequency using any SF without further configuration. When listening on the same frequency, only one receive path "locks" on the incoming signal, leaving the others free to detect more incoming packets. If there are no available receive paths when a packet is to be detected, that packet is lost [21].

4.1.3 Channel Line-up and Application Model

The original module assumed operation for devices in Class A only. The channel access is ALOHA and is divided as shown in Table 4.3. Each node will choose one frequency channel at random for each transmission, provided it is available. The gateway receive paths are then distributed equally among them to maximise coverage on each frequency [161].

Devices are configured to generate periodic uplink messages, according to [164]. No downlink transmissions are considered and to avoid synchronisation, each end device is allocated a random initial delay, following which the node starts with the periodic routine.

Table 4.2: Packet rejection matrix in the original NS-3 model [161], from [71]

	SF7	SF8	SF9	SF10	SF11	SF12
SF7	-6	16	18	19	19	20
SF8	24	-6	20	22	22	22
SF9	27	27	-6	23	25	25
SF10	30	30	30	-6	26	28
SF11	33	33	33	33	-6	29
SF12	36	36	36	36	36	-6

Table 4.3: Channel line-up from the original NS-3 module [161]

Channel Index	Centre Frequency (MHz)	Receive Paths
1	868.1	3
2	868.3	3
3	868.5	2

4.2 Improving the Starting Model

The original model has, since its creation, gone through a number of upgrades and enhancements which currently place it amongst the most complete ways to perform discrete-time simulations for LoRa and LoRaWAN [72]. Some of these changes have now been incorporated into the downloadable models, freely available at [162] as part of the community's effort to update them along with their creators. This includes a number of modifications performed by this thesis's author, as documented in a number of GitHub issues [167–170]. Some others are currently only available as part of this thesis's author module, available at [171]. In this section, the main contributions and upgrades will be introduced and compared to the behaviour of the original module described in the previous section, when appropriate.

A few changes were made to the way the simulation creates packets, after noticing that the Time-on-Air of simulated packets was slightly different from that obtained by popular LoRa ToA calculators [172, 173]. It was first noticed that the Low DR Optimisation bit, which is mandated to be set to 1 for T_{sym} greater than 16 ms [49], was always set to 0. This should not be the case, as it should be set to 1 for SF11 and SF12 operating with a BW of 125 kHz, which give a T_{sym} of 16.3 and 32.7 ms respectively, using Equation (2.1). This change is now included in the downloadable models, and documented in a GitHub issue [168].

Even with this change, the ToA was still slightly off, with 4 more bytes needed to be added to the length of a packet to make it congruent in ToA with what the calculator suggested [172, 173]. It was then noticed that throughout the models there is no mention to the 4-octet Message Integrity Code (MIC) [174] that should be added to the header according to the documentation [41]. An issue was created recently and is currently open [170]. In the models used in this thesis, this change was made manually at the start of the research work by simply adding 4 bytes to the PHY header. Thanks to the fix in the data rate optimisation bit and the addition of the 4 bytes MIC, the ToA obtained through NS-3 simulation was made to match with that of online calculators, with no observed detriment to the simulation [172, 173]. This is currently only featured in this thesis author's module [171].

The channel line-up was also enhanced, from only featuring the standard, mandatory

frequencies of 868.1 868.3 and 868.5 as shown in Tab. 4.3, to also support the full 8 channels which are part of the EU frequency plan, such as the one used by The Things Network [57]. The sensitivity of devices was also updated, from the original, shown in Tab. 4.1 to that obtained from the documentation available at [44, 46] and reported (for gateways), in Tab. 2.1. These changes are currently only available in this thesis author's module at [171].

Changes were made to how the module dealt with interference. Firstly the matrix modelling pure ALOHA behaviour was fixed, as it was not losing all packets involved in case of a collision. This is documented in a GitHub issue [167] and now part of the original module.

A brand-new collision matrix was also added, to reflect the new and improved version developed by [39] over that reported by [71] (see Tab. 2.4), used in the original module. This change only applies to the author's module at [171].

The behaviour of a transmitting node was also updated. In the original module, a node that was found trying to transmit over its Duty Cycle limit would queue the same transmission to send again as soon as a channel opened again for transmission. This is not how the RN2483 LoRaWAN chip [68], used throughout this thesis, operates. In the case of this LoRa transceiver, the same transmission would not be queued up but lost instead, with the device moving on to the next one. This change, provided by simply commenting a line of code in charge of rescheduling a stopped transmission, is only available in this thesis's author module [171].

A more considerable and sizeable enhancement is the one made to the application layer. In the original module, each device could only send either a single packet, or a number of them periodically, with a fixed period and parameters (which were only changeable by the ADR algorithm). More complex transmission behaviours were implemented, where different nodes can have different schedules, change SF and period as they go without requiring downlink. This was key for the work carried out in Chapter 6, and currently only available in this thesis's authors module [171].

A shared effort was instead at the base of an important update as to how the packet reception works at a very fundamental level. In the original module, as shown in Table 4.3, each frequency channel was assigned to a single frequency, and could only receive packets sent on that frequency. By going through section 3.6 of the documentation on the SX1301 [74], it became clear that this was not the case. Instead, demodulation channels do not link directly to any one frequency but can detect preambles and latch on to them on any frequency and any SF, meaning that a gateway could technically receive 8 simultaneous packets sent on the same frequency, rather than a maximum of 1, if each demodulation path was assigned a single frequency. A consequence of this is that the likelihood of a packet being lost due to a lack of demodulators decreases. This update is now available

as part of the original modules, and documented in a GitHub issue [169].

Finally, the whole data logging that was present in the original module was enhanced. From simply getting information about the network's PDR, in this thesis author's module [171] more granular information can be recorded including: information about packets attempted and sent by each device, packets outcomes per device both at the gateway and at the network server level, total traffic generated and recorded both in uplink and downlink by each device, interference information and RSSI information. This data is collected in text files throughout the simulation run, allowing further processing and various metrics extraction to be done through MATLAB, post simulation.

4.3 Simulation, Data Analysis and Other Tools

This section includes details of other software and tools that were used during the course of the PhD to help with the set-up as well as streamlining the execution of many simulations, recover the data from the LoRaWAN NS, analyse it and plot it. An overview of the tools and their usage is also included in Table 4.4.

4.3.1 SEM

A python simulation manager for NS-3, Simulation Execution Manager (SEM), also developed by the same researchers as the original LoRaWAN module, and available online [175, 176], was used. This makes for faster and more flexible simulations by allowing multiple NS-3 runs of the same script, with configurable set of parameters, as well as results management, and data collection. Rather than having to initiate every simulation and manually altering parameters such as numbers of nodes, or simulation time, one can simply set these to be performed in all combinations of parameters via a Python script.

4.3.2 MATLAB

The data analysis for the work undergone throughout this thesis is carried out almost entirely on MATLAB, with only occasional integration of Microsoft Excel, for the sake

Table 4.4: Overview of the software tools used

Tool	Usage
NS-3	Setting and running discrete-time simulations of LoRaWAN networks
SEM	Automates batch testing of the same NS-3 with varying parameters
MATLAB	Data Analysis and discrete-time simulator for the work performed in [27]
Network Server	Collects and passes on packets which are linked to our IoT applications
Node-Red	MQTT endpoint for the packets arriving at the Network server

of ease and speed. In the MATLAB scripts developed, text files generated by NS-3 and SEM are loaded in, parsed, and the data is then processed in specific ways to the work to be performed. These include plotting how many packets each gateway received in a multi-gateway network, as well as creating bar plots showcasing the percentage of packets sent by a node that were lost to each different outage condition. Additionally, to perform the work contained in Section 5.2 MATLAB was also used as a discrete-time simulator by creating a script mimicking the basic LoRa and LoRaWAN functionalities.

4.3.3 Network Server

Initially, the choice of network server naturally fell on The Things Network, which would then eventually become The Things Stack [177]. At the time, this was the most complete source of LoRaWAN open-access information, as well as one of the providers of free-to-use, out-of-the-box solutions for network access. Among these, the integration with Hypertext Transfer Protocol (HTTP) requests and Message Queue Telemetry Transport (MQTT) proved fundamental in gathering the data arriving from the deployed end devices. The Things Products such as The Things UNO and The Things Gateway were also chosen for ensuring maximum compatibility. The network server and tools provided by The Things Network and The Things Stack were used throughout the PhD and were only superseded by the network deployed by IoT Scotland [178] for the work presented in Chapter 6. IoT Scotland's network was launched in 2018 and offered many of the same advantages, plus an increase in coverage, currently connecting 29 of Scotland's 35 Local Authority areas, 35% of businesses and 1.9 million people [178]. While the two networks should operate in an identical manner due to regulations and restrictions imposed on LoRa and LoRaWAN in EU, and therefore not be reason for any discrepancies in results, this was not formally tested and left as future work.

4.3.4 Node-RED

Finally, Node-RED was used as an end-point for MQTT communication coming from the NS. In Fig. 4.1 a typical signal chain can be seen. Payload and metadata packets coming from the NS via an MQTT broker are converted to JSON and parsed through JavaScript to only collect relevant information, which is then saved to text files and, when required, plotted onto a custom dashboard.

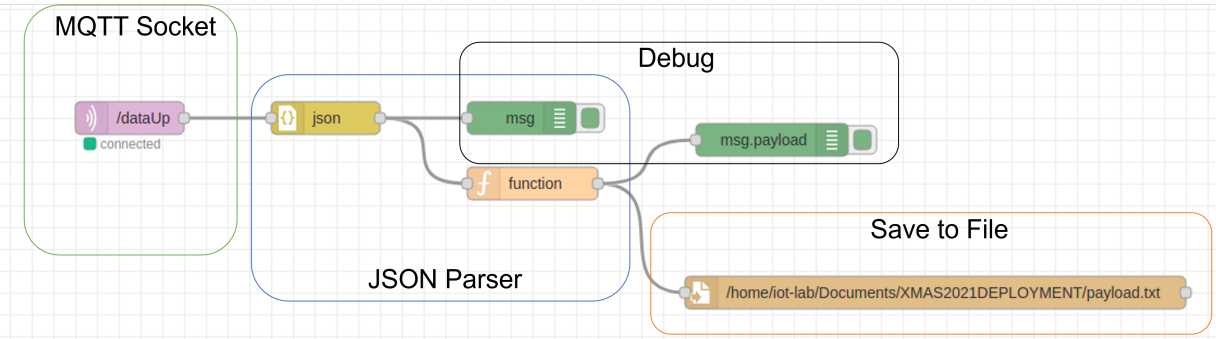


Figure 4.1: The integration of Node-RED, saving all data and metadata in text form

4.4 NS-3 Comparison and Validation Against Stochastic Models

As mentioned in Section 3.2, both stochastic geometry and discrete-time simulations are effective methods for analysing otherwise impractical large-scale LoRaWAN networks. Because both techniques are utilised in different circumstances in the literature, it is critical to prove that the results from the two distinct analytical methods are similar so that they may be used interchangeably depending on the simplicity of application in the given context. This gave us, additionally, the chance to validate the updating work performed on the original NS-3 modules. In this section, published originally as a Conference paper [27], such a comparison is performed by replicating in NS-3 a common result from stochastic analysis of a single gateway network. Results of the comparison show that the modules we enhanced are working as intended and that the two analysis methods are equivalent between 90 and 95% accuracy, and thus they are both equally employable in future research.

4.4.1 NS-3 Simulation Setup

Stochastic models typically rely on a few assumptions which effectively "downgrade" LoRaWAN's behaviour, to keep the complexity of the resulting mathematical analysis low. The LoRaWAN modules developed and described in Section 4.2 went through additional changes specific to this work to replicate at the best of our possibilities the assumptions and setup of the stochastic analysis performed. These were gathered by analysing three different papers [17, 143, 144] and also integrated with feedback obtained with direct correspondence with the paper's main author. Some of these changes to the NS-3 modules performed specifically for this comparison are:

- The usage of a single frequency channel at 868.1MHz, as opposed to the three default channels for LoRaWAN in the EU (868.1, 868.3, and 868.5MHz) or the eight utilised

by commercially available devices (868.1, 868.3 and 868.5, 867.1, 867.3, 867.5, 867.7, 867.9MHz)

- The implementation of a purely theoretical, infinite amount of parallel demodulation paths on a receiver, allowing for any number of packets to be received concurrently. In commercially available gateways, as in the unmodified modules, the number of demodulation paths is 8.

- The usage of a custom packet rejection matrix to deal with interference which specifies a 1 dB power difference for the capture effect to occur, while also assuming perfect orthogonality between different SFs, as opposed to the matrix reported by [71] or [39].

The number of nodes in the simulation is set to the average result of the Poisson Point Process using the intensity of 3 devices per km^2 to be uniformly distributed within a radius of 10755 metres. The SFs for the nodes are assigned using the standard distance-based technique of choosing the lowest feasible SF per node that will provide a strong enough link to be received after the resulting path loss. To match the one described in [143], the conventional Friis propagation model included in NS-3 was also adjusted to

$$PL^{dB} = (\psi/4\pi d)^{\gamma}\ dB \tag{4.6}$$

where:
ψ is the wavelength of the carrier frequency in cm;
γ is the path loss exponent, in this case 2.7;
d is the link distance;

The idea behind this spreading factor allocation, as well as the formulae needed to obtain the following boundaries will be explained in further detail when discussing more in-depth research in the next chapter. In brief, by knowing the required sensitivity of the receiver for each spreading factor, one can find the highest d that gives a sufficiently high sensitivity per SF in Equation (4.6). This enabled to find consistent SF boundaries with those in [143] (3259, 4209, 5436, 7021, 8690, and 10755 metres). Combined with the distance-based SF allocation mechanism, this resulted in the network topology illustrated in Fig. 4.2.

A change that was not possibly replicable was the specific transmitting schedule of the nodes. In their work, the authors of [17] specify a shorter packet ToA than what should be achievable using LoRa and due to the complexity of changing the underlying modules' operation this was deemed non-replicable. In our simulation instead, each node transmits 12 bytes of payload (plus 13 of necessary overhead for LoRaWAN) at a fixed

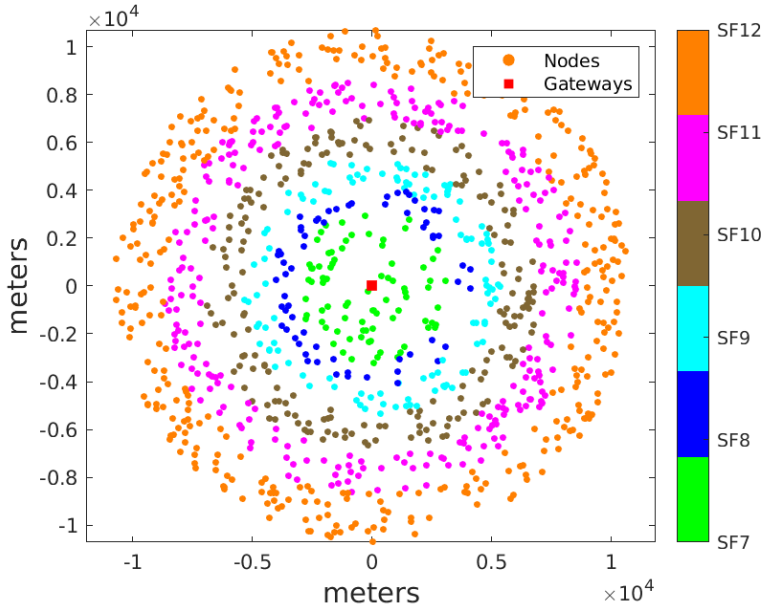


Figure 4.2: Network topology featuring a single, central gateway and uniformly distributed nodes with SF allocation based on their distance to the gateway

speed that is dependent on the node's SF and the ToA of each packet. Each node hence transmits as often as it can while remaining under the duty cycle limitation of 1%. This gives transmission delays of 6.138, 11.286, 20.394, 40.788, 81.576 and 146.817 seconds for SFs from 7 to 12, respectively.

Additionally, the authors include an element of randomness to the packet schedule, with packets transmitting around a fixed time, so that on average the duty cycle limit is respected [143]. As this is not how LoRaWAN chips such as the RN2483 [68] operate, this also was not implemented. More complex transmitting behaviours including similar randomness will nevertheless be designed for further work described in this thesis.

4.4.2 Results & Analysis

The particular result that was recreated is the probability of packet loss as a function of the distance of nodes to a single, central gateway. This result, with only minor differences, is found in a number of papers such as the ones the simulation is based upon and also work performed by other researchers [19].

While the details and setups of the models in these works are not the same, the resulting graphs are comparable. The calculated, theoretical PDR, which in this case is equivalent to the probability that a packet survives all outage conditions brought forward, has in all cases a distinctive "sawtooth" shape, as shown in Fig. 4.3.

Visual analysis of this probability raises a couple of important points. As expected, the overall chance of a packet being received and decoded correctly drops considerably the

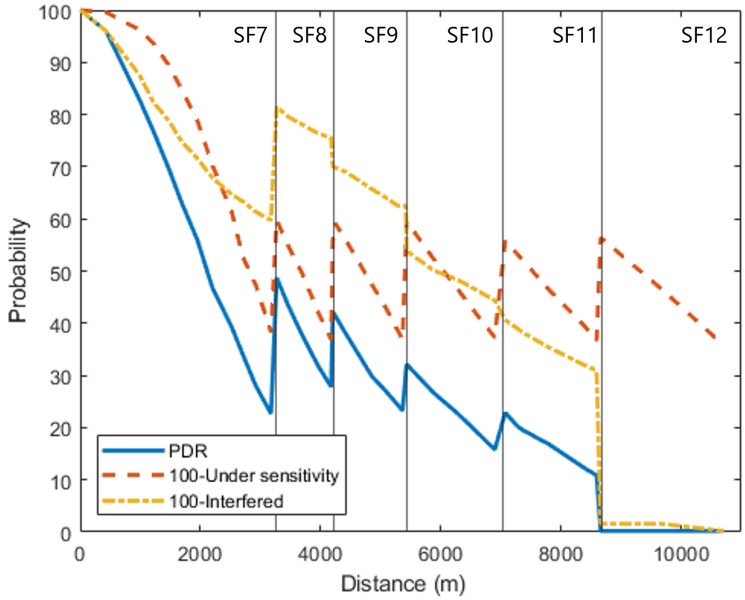


Figure 4.3: The success probability of a packet against the link distance between node and gateway, according to the stochastic geometry models from [143]. In blue the overall success probability, in orange the complementary to the probability of collision and in red the complementary to the probability of the packet being under sensitivity

further from a gateway the transmitting node is. The link deteriorates with increasing distance and increasing signal path. The distinctive "sawtooth" shape of the red and blue curves in Fig. 4.3 is the probability that a packet survives the under sensitivity outage condition and the overall success probability of a packet, respectively. This shape is due to the fact that once a certain link distance d is reached, a node will have to switch to the next SF to lower the threshold in receive power that the gateway needs to decode its packets correctly. This causes an immediate jump in the chance that a packet won't be found under sensitivity after factoring time-varying effects such as fast fading. This is then followed by a progressive decline of this probability, as the node moves further away within the same SF region, thus increasing its PL, until a new boundary is reached. The declining chance of passing the interference check, in orange, is also explainable by reflecting on the peculiar characteristics of LoRa, namely, the capture effect. As a node is closer to the lower boundary of each SF, hence to the gateway, its SNR and RSSI at the receiver should be higher, as the signal has a shorter path to the receiver. This means that when packets being transmitted with the same SF collide, those coming from nodes which are closer to the receiving gateway will generally have a better chance of surviving the capture effect as they will arrive at said received with a better RSSI. This is also known as the near-far problem [147, 151].

As a first step towards the comparison and validation of the NS-3 models, an attempt was made to recreate this same result using the discrete-time simulation offered by NS-

3, after matching inputs to the best of possibilities as described in Section 4.4.1. The resulting graph can be seen in Fig. 4.4.

While the probability of a packet to be found under sensitivity and the overall PDR curves look satisfactory given the inherently different methods used to recreate them, the probability of packet collision and interference, in orange (Fig. 4.4) was very different, dropping at each SF boundary but steadily increasing the further away from a gateway.

After consideration, this was deemed to be because of the way the two causes of loss are treated independently in stochastic models and calculated regardless of one another. In reality and in the NS-3 simulation, this is not the case. A packet needs to be first "locked" on by a gateway by having a high enough power, hence surviving the sensitivity check. Only then it can collide and possibly fail because of interference with others. Therefore, while not all packets are checked for interference and collision, in case they already failed the sensitivity check, their power will always add up to the interference present on that frequency and on that SF. Because fixing this discrepancy would have meant uprooting the whole NS-3 ecosystem and a deep rewrite of not only the LoRaWAN module, but also some native functionalities, a compromise was reached. This involves acknowledging that the number of packets that collide and interfere is not a percentage of all packets that have been sent, as these include a part that may be lost to path loss, but instead, a percentage of the number of packets that have already successfully passed the sensitivity check. By slightly modifying the plotting functions to take this dependency of one outage condition upon the others, rather than treat results independently, Fig. 4.5 is reached, which is, this time, very similar to the original results for both probabilities and their sum.

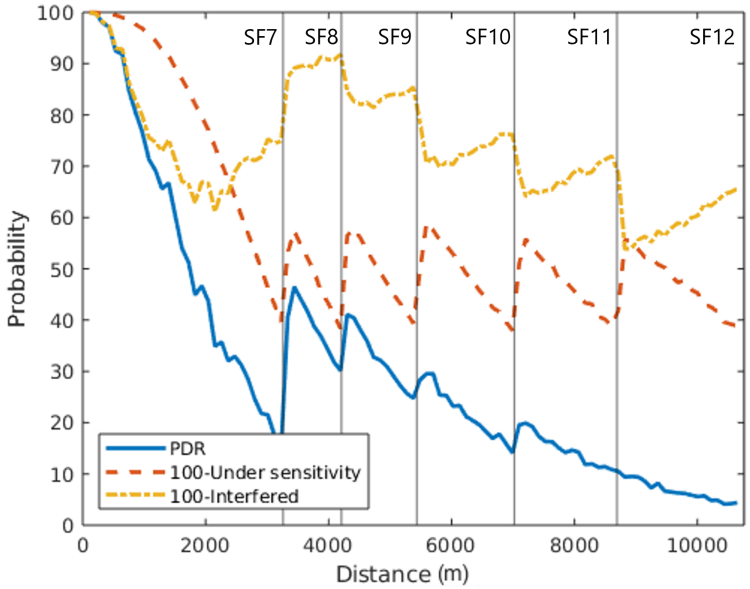


Figure 4.4: Initial result for the PDR (blue) and the complementary to the outage probabilities as a function of a node's distance from a central gateway

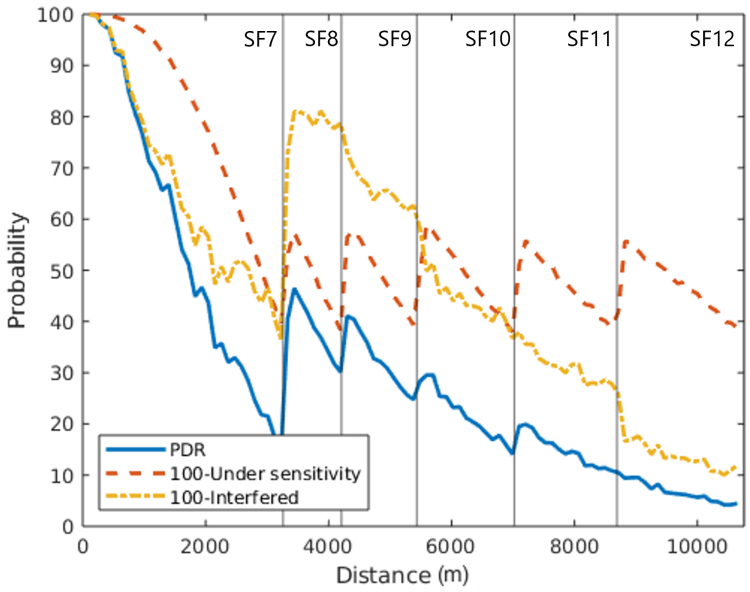


Figure 4.5: PDR (blue) and complementary to the outage probabilities as a function of a node's distance from a gateway by taking into account timing discrepancies between the two analysis methods

Alongside a small discrepancy of roughly 3-5%, the main appreciable difference is in the shape of the interference chance as well as the performance at SF12. Despite the network is shown to perform roughly 5-10% better in simulation than in the model while using the highest possible SF, we believe this error is tolerable when comparing the two different methods. This is especially the case considering that some assumptions of the stochastic setup, such as the shorter ToA and the random time interval between packets were not perfectly replicable in NS-3.

4.4.3 Summary and Key Findings

Despite their differences, this study demonstrated that stochastic geometry models and discrete-time simulations are interchangeable methods for analysing the performance of a LoRaWAN network, especially when high density and high scale is a focus of the tests. The comparison of the same network gives rise to results that are within a maximum on 10% of each other, which is acceptable given the exact same settings of the theoretical model could not be replicated in NS-3. Still, this is the first example and closest result in academic research, with no comparable works available at the time of writing. This similarity can be achieved provided that some significant distinctions between the two methodologies, in terms of the interdependence of components of protocol behaviour, are well understood and accounted for. Stochastic models have the benefit of allowing for a more in-depth view of the network's behaviour, with mathematical formulations to fully characterise it. On the other hand, discrete-time simulations are more of a black-box

approach, yielding the best results overall, but without giving as much insight on all the moving parts of a network.

This same study also proved that the modification that were being made to the NS-3 models throughout the course of this PhD were valid and operating as they should. Furthermore, new, additional enhancements for the models were inspired by it, such as the need for more complex node behaviour than simple periodic uplink. The validation of the modules performed here, as well as the expansions and improvements it inspired, would become the foundation for the published work carried out during the rest of the PhD and reported in the next two chapters of this thesis.

Chapter 5

LoRaWAN Scalability Analysis

Having established the NS-3 modules were performing as expected by industry standards, the next step would be to use these modules for studying LoRa and LoRaWAN behaviour, particularly in large-scale networks. To this end, the following chapter contains novel research carried out in accordance to the gaps in literature highlighted in Section 3.3. This research analyses different aspects of the scalability of LoRaWAN in the context of various network setups and has been published and presented in a journal [25] and at a conference [27], respectively.

5.1 Impact of Inter-gateway Distance on Network Performance

The intuitive idea that LoRaWAN performance, scalability and overall QoS can be boosted by increasing the number of gateways and their density is established [20, 21, 72]. In fact, part of the appeal of LoRa as a technology is the low capital cost and wider availability for users to set up gateways, benefiting the community as a whole. However, contrary to other technologies, because of the characteristics of LoRa such as the capture effect, the benefit additional gateways provide is widely dependent on their relative positioning.

The work contained in this section, and originally published in [25], aimed to test this assumption by performing an analysis of a two-gateway network.

5.1.1 System Model and Simulation Setup

The simulation is performed in NS-3, using the Log-Distance path loss model to approximate the power losses experienced by the wireless packets in an urban environment. The parameters for γ, PL_0, and d_0 are the ones used by Magrin et al. [21] and are reported in Table 5.1.

Table 5.1: Simulation parameters

Payload Length	10 bytes
CR	4/5
BW	125 kHz
Channel	868.3 MHz
Tx Power	14 dBm
Sim. Time	3600 s
$Tx_{interval}$	180 s
PL_0	7.7 dB
d_0	1 m
γ	3.76
Noise Figure (NF)	6 dB
Nodes (N)	500, 1000, 2000, 5000
Radius (R)	3011, 3509, 4089, 4766, 5554, 6473 m
Distance between GWs (D)	0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1
Simulation Hardware	11th Gen Intel Core i7 @ 3GHz

To calculate the maximum distance a node can be placed from any gateway, time-based losses are not considered. Instead, only the relative distance between a node and a gateway is used to calculate the overall power loss, in this case using the Log-Distance model. By inserting the parameters into Equation (2.13) and solving for d, the maximum link distance a signal of a given SF can be successfully decoded at can be found with the following:

$$d \leq 10^{\frac{P_t - PL_0^{dB} - X_g}{10\gamma}} d_0 \; metres \tag{5.1}$$

The results for the different gateway sensitivities for each SF, using the parameters found in Table 5.1, are in Table 5.2.

The Log-Distance PL plus time-based losses, along with the total interference due to inter- and intra-SF collisions are used concurrently to determine whether a packet is successfully delivered and decoded at any gateway. The payload of each packet is set to 10 bytes, which, added with the 13 bytes of mandatory LoRaWAN overhead, makes for a SF12 transmission air-time of 1482.8 ms, as per Equation (2.4). The time between transmissions, $Tx_{interval}$, is set for all nodes to 180 seconds. This allows for the 1% duty cycle restrictions to be satisfied for any SF, according to Equation (2.10). The overall simulation time is set so that each node will send a total of 20 packets throughout the simulation run. The SF is allocated based on a node's distance from any gateway, as it is common. Each node is assigned the lowest spreading factor that can achieve a good link with at least one gateway. No ADR algorithm is applied, meaning that the transmission parameters of each node will not be changed throughout the simulation. In addition, no downlink transmissions or re-transmission algorithms are taken into consideration. The rejection matrix initially proposed by [71] is used to check whether packets survive interference

Table 5.2: Parameters for a network with a single gateway placed at the origin, using the log-distance path loss model, ignoring shadowing and fading

SF	**SNR_{min} @ Rx (dB)** [a]	**S_{GW} (dB)** [b]	**R_{max} (m)** [c]	**AoD (km^2)** [d]
7	−7.5	−124.5	3011	28
8	−10	−127	3509	10.2
9	−12.5	−129.5	4089	13.84
10	−15	−132	4766	18.83
11	−17.5	−134.5	5554	25.54
12	−20	−137	6473	34.72

[a] Minimum SNR at the receiver, from Reference [49]. [b] Gateway Sensitivity, using Equation (2.8). [c] Maximum distance from a gateway that a node at each SF can be placed at, using Equation (5.1). [d]Area of deployment for nodes operating on each SF, based on R_{max}.

from collisions. Finally, to model the hardware limitation of having a finite amount of demodulation paths on a single receiver, each gateway can "lock on" to, at most, 8 packets at a time.

The NS-3 simulation is then performed by varying the total number of end-devices $N = \{500, 1000, 2000, 5000\}$, the radius these are uniformly distributed on $R = \{3011, 3509, 4089, 4766, 5554, 6473\}$ metres and the distance between the two gateways in the network as a factor of the network radius $D = \{0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1\}$, with 0 representing the centre of the circular network and 1 the whole radius length on either side of the origin. For example, for $D = 0.5$, the two gateways will be placed at $0.5 \times R$ on either side of the network centre, giving an inter-gateway distance of R as shown in Fig. 5.1.

Values for D were chosen to give a good spread of possible inter-gateway distances, as was the total amount of nodes (N), which reflect typical values for network deployments found in literature [19, 132].

Values for the network radius (R) were chosen to be as close as possible to the calculated "boundaries" between the different SFs a node can adopt. For instance, every node transmitting in a network with a single central gateway, operating under the parameters configured and before applying time-dependant power losses, will be able to operate at SF7 if the overall network has radius $R = 3011$ m. However, if the same network had $R = 6473$ m, every SF will have to be used, with SF8 being used by devices in the first annulus, between 3011 m $< R <$ 3509 m, and so on.

This is complicated by the addition of a second gateway. Under the assumptions and parameters presented in this work, the maximum distance that a single gateway placed at the origin of a network can cover is a circle of radius 6473 m around itself. Based on this, it can be demonstrated that it is impossible for two different gateways to be placed anywhere other than the network origin while covering the entire deployment area in a network with nodes uniformly distributed over a radius of 6473 m. This is illustrated in

Fig. 5.2.

The maximum distance two separate gateways can be placed at to cover the entirety of the uniformly distributed nodes can be calculated with the following:

$$z = \sqrt{\frac{R_{max}^2 - R_d^2}{R_d^2}}\ metres \tag{5.2}$$

with:
R_{max} = maximum distance a node can be from a gateway on a given SF, from Table 5.2;
R_d = radius of the network deployment.

Results for z for the different network radii are given in Table 5.3. Given this, only simulations with $D \leq z$, to ensure that every node is within coverage, are performed. As an example, for simulations with $R = 6473$ m, only $D = 0$ is taken into consideration; for those with $R = 4766$ m, every value except for $D = 1$ can be used.

The simulation is first performed as described, specifically ignoring any time-based power loss effects, and then performed again, this time to include fast fading to the Log-Distance model. This is modelled using the Nakagami-m model with m = 1, to equate the Rayleigh distribution [179], and concatenated with the existing path loss.

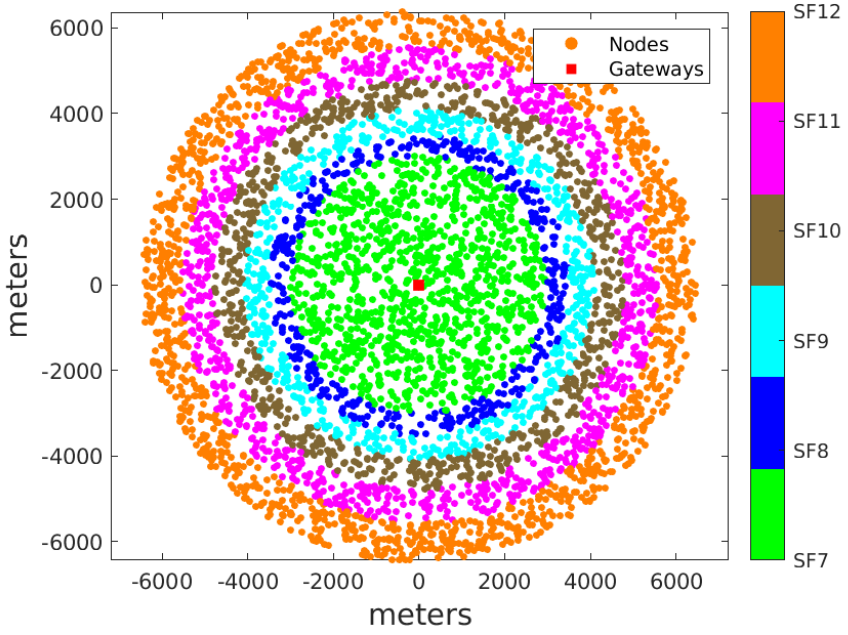
The results of the simulations, obtained with NS-3, were further analysed and plotted through dedicated MATLAB scripts.

5.1.2 Simulation Results

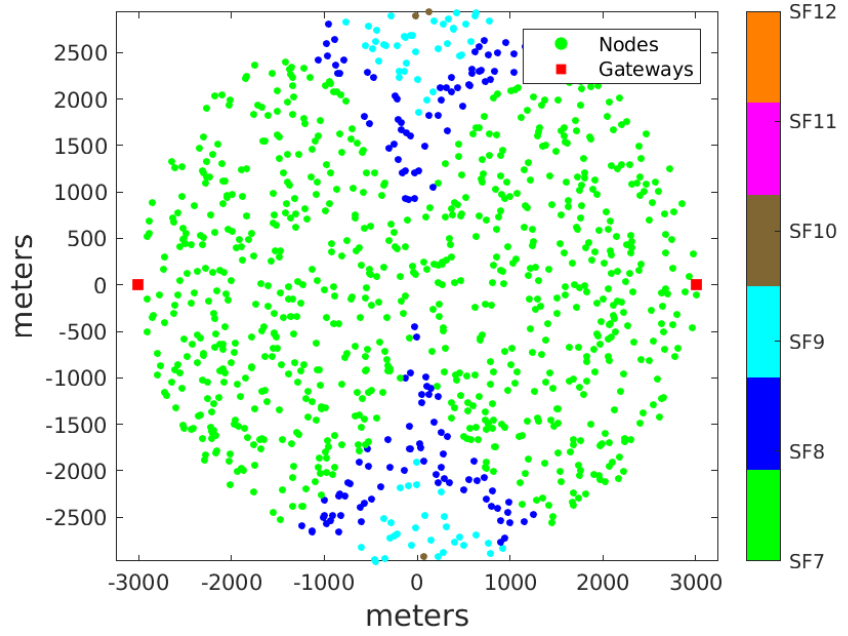
Results in Fig. 5.3 show that increasing the number of nodes, as well as increasing the network radius has the general effect of lowering the overall PDR. Nodes in these cases have to switch to higher SFs to ensure connection with at least one gateway, thus increasing the chance of collision. This result is true across all possible combinations of R, N and D, and it is a documented limitation of LoRaWAN scalability [66]. More interestingly,

Table 5.3: Maximum distance a symmetrical pair of gateways can be from the centre of a deployment so that all devices will be contained within the combined coverage of the two gateways, z

SF	**R_{max} (m)**	**z**	**Range of Valid D for Full Coverage**
7	3011	1.9	0–1
8	3509	1.55	0–1
9	4089	1.22	0–1
10	4766	0.91	0–0.875
11	5554	0.59	0–0.5
12	6473	0	0

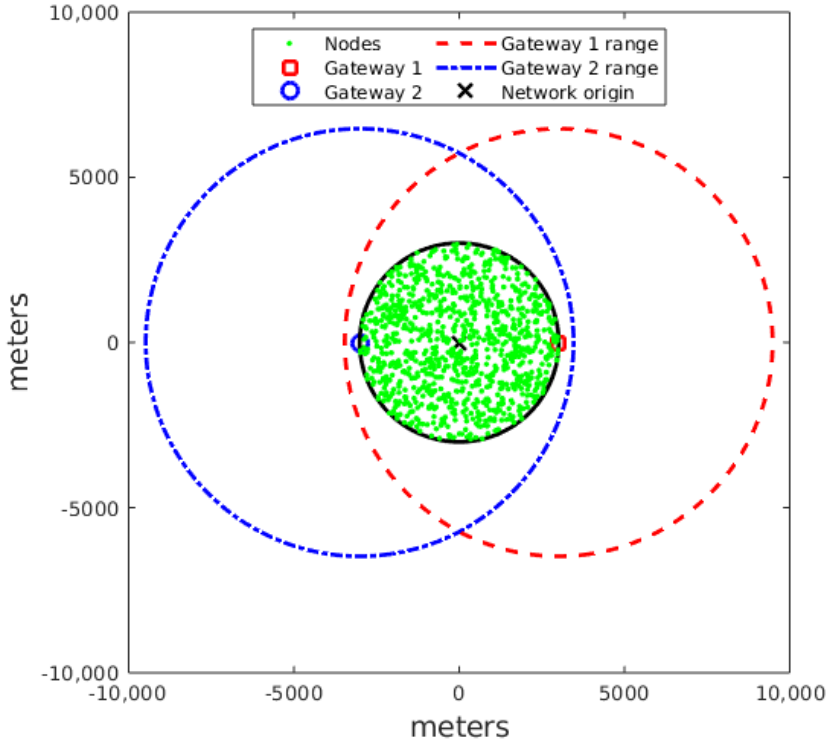


(a) Two gateways, $N = 5000$, $D = 0$, $R = 6473$ m

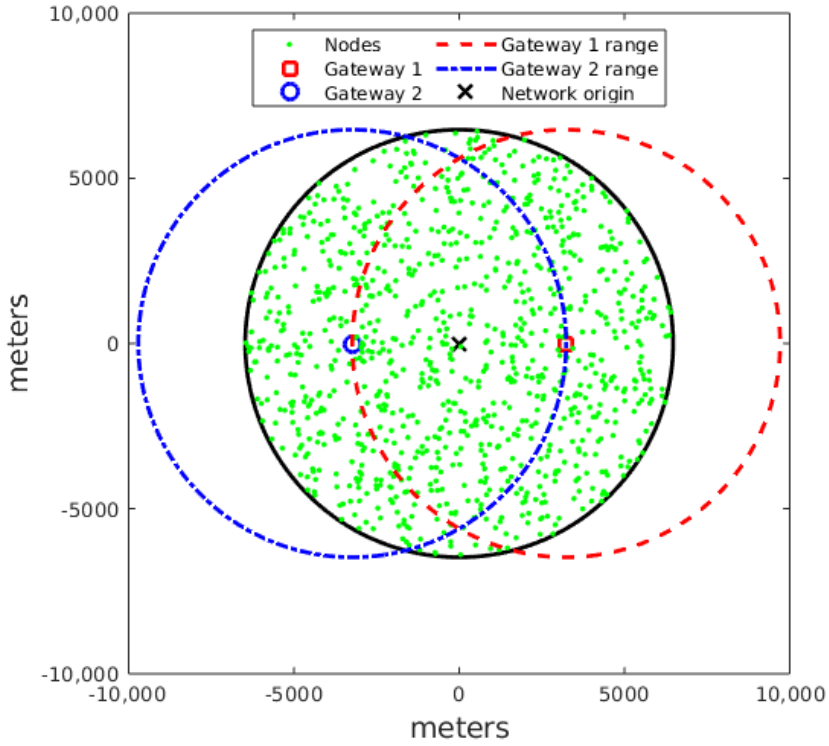


(b) Two gateways, $N = 2000$, $D = 1$, $R = 3011$ m

Figure 5.1: Two examples of network topologies, showcasing the different SFs distributions with varying network parameters N, D, and R



(a) $D = 1$, $R = 3011$ m. All nodes deployed are covered by both gateways



(b) $D = 0.5$, $R = 6473$ m. A portion of the deployed nodes is not covered by either gateway

Figure 5.2: Examples of different gateway distance (D) and network deployment radius (R, black circle) scenarios

the simulations show that increasing the distance between the two gateways generally increases the PDR.

As a result of the gateways moving away from the network's origin, more nodes will be able to transmit with a lower SF. Due to this, packets from some of the nodes with lowered SF will only be received by the closest of the two gateways, increasing the likelihood of interference and fast fading causing an outage, with no chance of the other gateway receiving and thus "saving" the data. The opposite effect can be seen for the collision probability. As D increases, this probability decreases. This is also because of the introduction of higher SFs in the network. In networks with devices operating using a single SF, the chance of collision will be highest. Moving the gateways apart, higher SFs have to be introduced, hence decreasing this probability, leveraging the SF quasi-orthogonality and the capture effect. This is visible in Fig. 5.1b, where moving the two gateways to $D = 1$ introduces devices using SF8, 9, and 10 to a network that, if both gateways were placed at the origin, would only include nodes operating with SF7.

The benefits of increasing gateway density are understood in literature for both LPWAN and, specifically, LoRaWAN, with a direct correlation between number of gateways deployed and PDR [16, 20, 21].

However, in a LoRaWAN network, this benefit is heavily dependent on the position of the gateway deployed. This is mostly due to specific features of LoRa and LoRaWAN, such as a star-of-stars network topology and the capture effect. In an unrealistic scenario where two identical gateways with identical antennas were to be placed in the exact same location ($D = 0$), the second gateway would contribute nothing towards the PDR of the network, unless the first gateway had failed [41]. In standard LoRaWAN, packets delivered to multiple gateways are de-duplicated by the NS and have no inherent use except redundancy and providing the network server with additional information to perform better ADR [77]. This is clear from Fig. 5.3, where lower distances between gateways, D, can be seen to generally yield lower PDR relative to the same configuration with higher values of D. Similarly, a drop in PDR can also be noticed as the values for D increase towards the maximum of 1 in most configurations, making values of D around 0.5 generally yield the best PDR. This reduction in PDR when further distancing gateways may be explained by considering the impact of this distance on all potential outage scenarios, as illustrated in Table 5.4.

Channel saturation is only seen to become an issue for networks when a great number of nodes ($\geq$2000) is involved, and the network has a radius such that higher SFs ($\geq$11) are needed. In Fig. 5.4b, with a 6473 m radius and more than 2000 nodes, it is visible that this particular type of packet loss rises significantly more sharply than the others as the network scales up. In the circumstance of this study, it appears that the inter-gateway distance can be increased as long as the resulting distance between a node and either

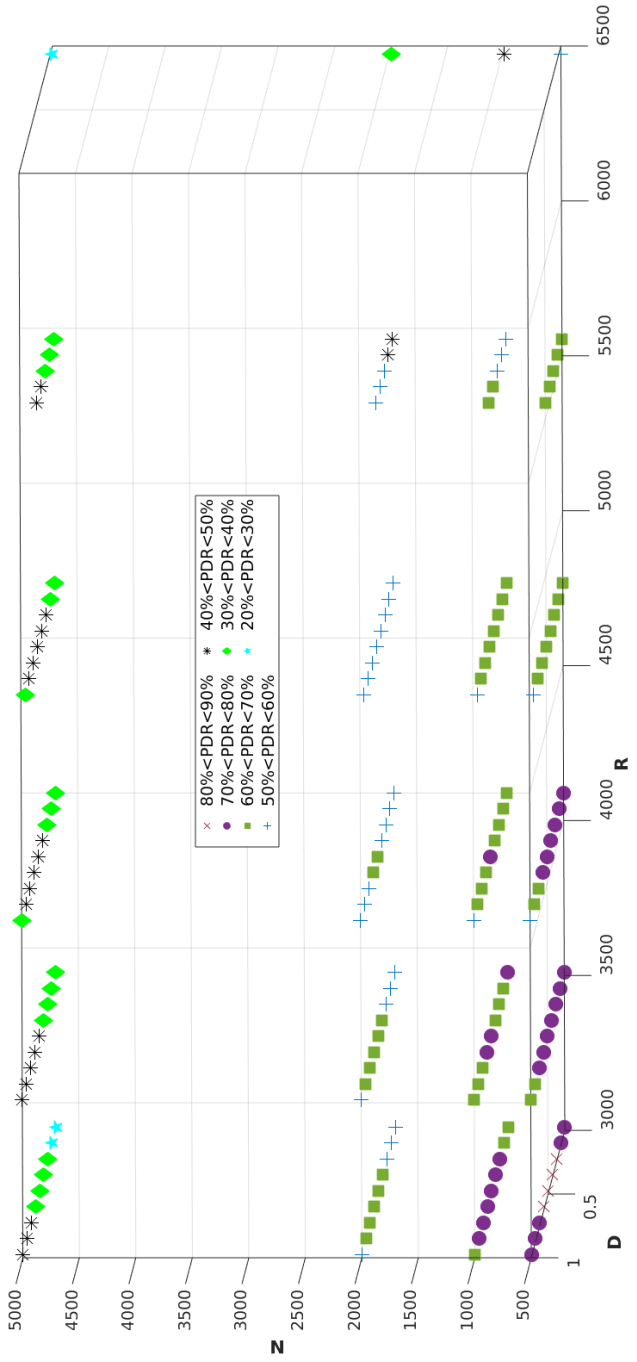


Figure 5.3: Three-dimensional scatter plot for the PDR across all simulations with fast fading enabled performed in this study, highlighting how low D simulations have lower PDR compared with higher D with otherwise identical configuration

Table 5.4: Breakdown of the weight of different outage conditions on a selected range of network configurations, all with fast fading enabled

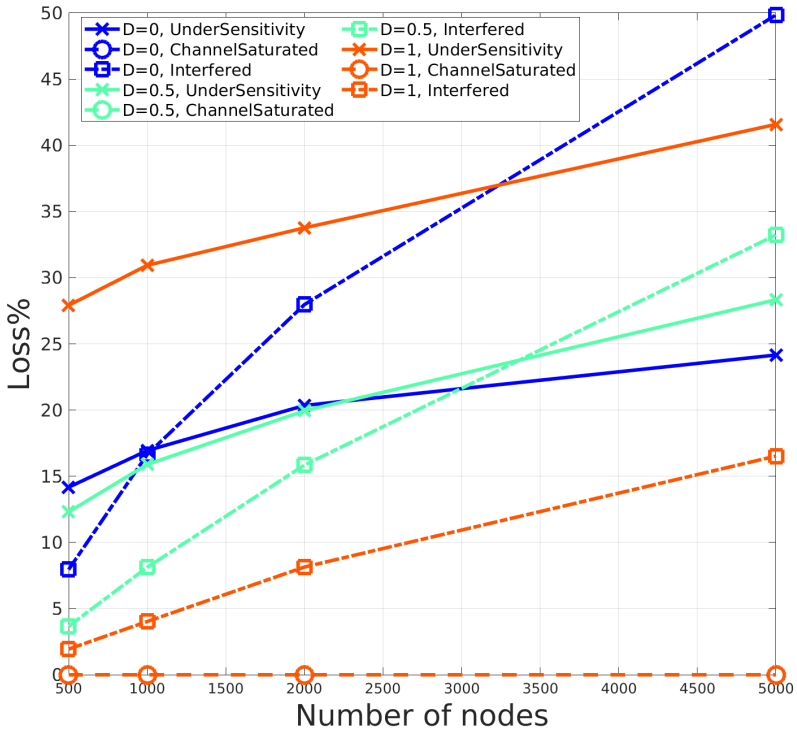
N	R	D	Sensitivity Loss%	Collision Loss%	Saturation Loss%
500	3011	0	14.17	7.95	0
500	3011	1	27.91	1.94	0
5000	3011	0	24.1605	49.8485	0
5000	3011	1	41.585	16.522	0
500	4089	0	21.395	4.255	0
500	4089	1	38.41	1.64	0
5000	4089	0	31.8035	31.291	0.0445
5000	4089	1	47.639	13.258	0.044
500	6473	0	31.825	10.605	0
5000	6473	0	30.697	23.1245	24.2625

gateway does not require a node to transmit at SF10 or higher.

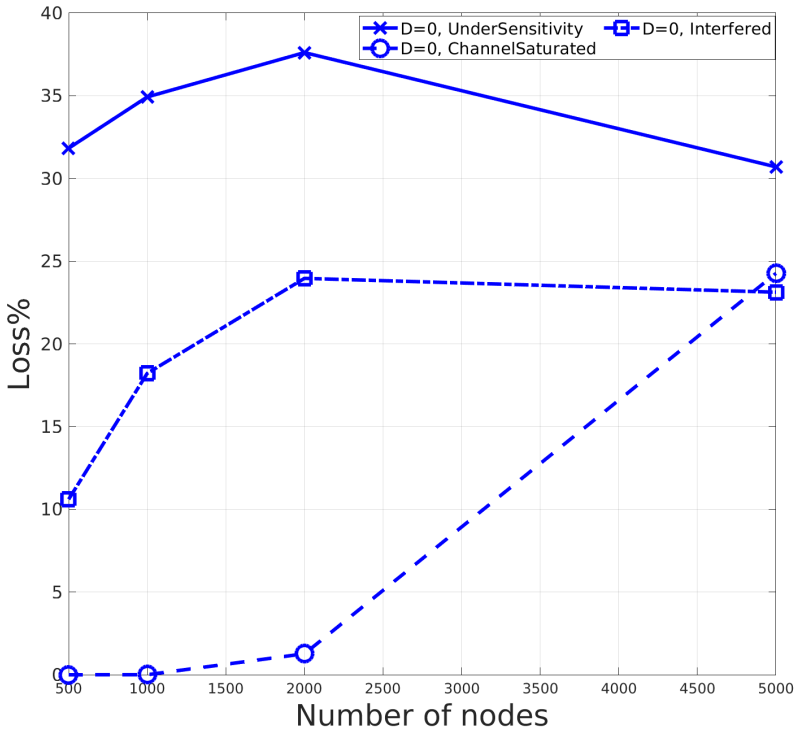
Overall, results in Table 5.4 and Fig. 5.4a show how the trade-off between the different outage conditions makes it beneficial to a network operating under the assumptions described in Section 5.1.1 and having low node density and low overall network radius to introduce nodes using higher SFs. However, if the network deviates from this arrangement, and especially as the network scales up its node density, this benefit soon disappears. For $R = 3011$ m, with $N = 5000$, increasing D from 0 to 1 increases the chance of packets being found under the minimum sensitivity required at a receiver by roughly 17% while also decreasing the collision probability by 35%, yielding an overall boost to the network's PDR for $D = 1$. For $R = 4089$ m and the same number of devices, $N = 5000$, increasing D from 0 to 1 has no additional benefit on the difference in percentage of packets being under sensitivity, which stays around 16%. However, now the collision probability does not decrease as much by increasing D from 0 to 1, with a smaller decrease of just 18% compared to the previous 35%.

The negative effects that introducing SF11 and 12 have on the PDR of similar networks are greater than the benefits they introduce by increasing the range of end devices and reducing the chance of same SF collision. This is because of the increase in the chances of time overlap between packets, possibly causing inter- and intra-SF collision, and the chance of a receiver's demodulators path saturation due to the elongated ToA. This contradicts the intuitive notion that creating the strongest feasible link between a node and a gateway is always advantageous, as also reported by [180].

Fig. 5.5 depicts the lines of best fit through all simulations performed for this work per SF. Increasing the total percentage of nodes in a network operating at any given SF is shown to be especially bad on the average network's PDR in the case of SF11 and 12. The increasing slope of the SF7 fit seems to suggest that, ideally, a network configured as



(a) $R = 3011$ m.



(b) $R = 6473$ m.

Figure 5.4: Total loss percentages for the different outage conditions across multiple N, with fast fading enabled

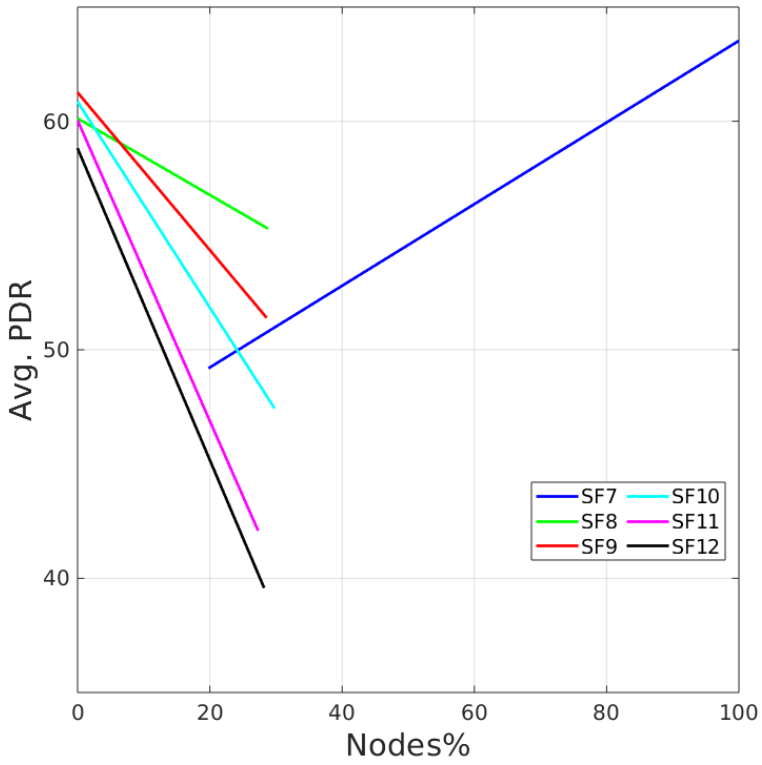


Figure 5.5: Average PDR of a whole network as a function of the percentage of nodes operating at each SF in the same network, with fast fading enabled

described in this work should have as many nodes operating at SF7 as possible. In the simulations where only 20% of the nodes deployed were operating with SF7, the average PDR was found to be just below 50%, while for those simulations with 80% of the nodes operating with SF7, the average PDR was 60%. Similarly, the average PDR when a simulation had no nodes operating with SF12 was just below 60%, while for those with about 27% of the nodes using SF12 the PDR declined to an average of 40%.

This is not always feasible, but nodes operating with SF11 and SF12 should be drastically limited in such networks, either by keeping the deployment size small enough that they are unnecessary, by increasing the number of gateways present, or implementing adequate and more complicated SFs assigning methods, as research highlighted in Section 3.2.5 suggests.

5.1.3 Summary and Key Findings

Taking as a reference a network with two gateways, a correlation between their relative distance and the network QoS was established via NS-3 simulations. This is not only regarding the network's overall PDR, but to the weight of each outage condition in an uplink-only scenario. While multi-gateways network are studied in literature [124,144,181],

no previous work at the time of writing went so deep into characterising the weight of the different outage conditions and how they related to the gateway positioning. The results show that, in a network operating in a similar fashion, with equivalent device density and distribution, similar and fixed periodic node transmissions, and the same SF allocation method, gateways are best placed at a distance that is roughly equal to the network radius, and equidistant from each other. For such a network it is shown to be advantageous to incorporate nodes with greater SFs, though as the network deviates from this configuration, particularly as the network builds up its node density, this is no longer the case. Starting from 2000 nodes, the downsides of introducing SF11 and SF12 outweigh the advantages they bring, as seen in Fig, 5.4. It follows that the SF allocation method based on spreading all the possible SFs across all nodes is detrimental for larger and denser networks operating under an equivalent setup as the one described in this study. In such networks, the goal should be to minimise as much as possible the amount of nodes transmitting with larger SFs. As we have shown in Fig. 5.5, the PDR of a similar network decreases rapidly as more devices operating at SF11 and 12 are added, with an approximate loss of 7% PDR for each 10% increment in the amount of nodes operating with those SFs.

5.2 Downlink Independent Scalability Optimisation

5.2.1 Background

A crucial aspect of the proposed enhancements to LoRaWAN scalability that were introduced in Section 3.2.5 is that they require some amount of downlink. Feedback from the network server is needed to instruct the nodes as to when to transmit their next packet and, for those who also employ dynamic parameter allocation, with what parameters. Ultimately, they require a considerable amount of planning and additional resources to what it would normally be needed to simply connect a node to a LoRaWAN network. One additional avenue of scalability enhancement not requiring downlink that has not been mentioned is re-transmission schemes. These are generally attempting to go around the lack of acknowledgement, which requires downlink, by simply transmitting the same packet multiple times during a single transmission interval [182, 183]. The aim is to maximise the chance that at least one transmission containing the data is received, and such schemes are usually employed in tandem with more involved and downlink heavy ones [131]. Another approach to context-unaware QoS improvement is diversity combining, as seen in [158], where many copies of a single frame can be recombined at the receiving end to try to recover the corrupted data. This is particularly useful in LoRaWAN, which, given its star-of-stars topology, is based on the assumption that multiple copies of the same packet will be received by all gateways in range. Normally these copies are de-duplicated by the NS and at most inform the ADR as to how to vary the duty cycle. Using diversity

combining, the network server could instead exploit this redundancy by recombining the received packet copies.

In this section we report work published and presented at the 1st IEEE Workshop on Sustainable and Intelligent Green Internet of Things for 6G and Beyond at IEEE-VTC SPRING 2022, for which it won the Best Paper Award [28]. Here, we analyse and test different ways to increase the quality of a network that would be easier and less resource demanding to implement in a real-world deployment. This is achieved by, rather than resorting to downlink feedback, only using the information available to the user on the end-device side.

5.2.2 System Model

A MATLAB script has been developed to recreate the transmission behaviour of LoRaWAN and explore the collision probability exclusively. This involves making sure no packet failure is possible for any reason but packet collision. Firstly, no node can ever be out of range, by selecting the size of the network deployment based on the maximum range of a device operating with SF12 and only subjected to the deterministic log-distance path loss as laid out by [21]. Similarly, the loss due to the limited number of parallel reception paths on a standard receiver is ignored, assuming that the gateway can decode an infinite number of packets simultaneously. This way, the Packet Delivery Ratio (PDR) can be calculated as the ratio of the packets that did not collide to the total amount sent:

$$PDR = \frac{not\ Collided}{sent} * 100\ \% \tag{5.3}$$

By knowing in advance the full radio parameters of all the nodes in the network at any given time, as well as assuring no other devices can be added and no external interference over the required frequencies, it would be trivial to create a schedule that ensures no packet will ever collide. With careful selection of SFs, channel frequencies and start times, such a network would yield a theoretical 0% collision rate even without downlink. As we are aiming to test methods that are as realistic as possible, as well as feasible to deploy in a real-world scenario, we will not consider this option. Instead, it was decided that the metrics and information the user and every single node can be aware of without the access to downlink transmission include the SF of the latest transmission, the transmission interval and the device start-up time.

The scenario was then split into two different use cases: one involves a full, proprietary network and another involves operating a single device which connects to an already existing network that is not known, akin to simply turning on a single node in an urban environment. The work performed intends to find ways to select the parameters that nodes are in control of to maximise the PDR of the whole network in the first case, and

of the lone device being operated in the second.

For both scenarios, the initial network behaviour is as follows.

A total of 1000 nodes are uniformly distributed around a central gateway, over a circular area with radius 6473 m. Propagation is modelled according to the Log-Distance propagation model.

A radius of 6473 m ensures that all devices will be within range with the parameters used, see Table 5.5. Spreading factors are allocated based on a node's distance from the gateway, with each device operating on a fixed SF which is the lowest it can possibly adopt. No downlink transmissions or uplink re-transmissions are taken into account. The network also operates in pure ALOHA fashion, hence ignoring the capture effect and the semi-orthogonality of the spreading factors.

Each node transmits 10 bytes of data with 13 bytes of mandatory LoRaWAN overhead, making the ToA of a SF12 packet equal to 1482.8 ms (Equation (2.4)).

The time between transmissions, hereby referred to as Tx Interval, is set for all devices to 300 seconds, to avoid issues due to the 1% duty cycle restrictions. Each node is set to send a total of 20 packets during a simulation run.

Each node is assigned a random start time which is between 0 and Tx Interval, in order to make sure that all devices will have performed 20 transmissions by the end of the simulation. For the simulation where the Tx Interval is modified, each device will only transmit a maximum of 20 packets.

Finally, each device will perform random channel hopping, selecting one among those that are free to transmit given the aforementioned duty cycle limitations.

Table 5.5: Simulation Parameters

Payload length	10 bytes
SF	7, 8, 9, 10, 11, 12
ToA	0.062, 0.114, 0.206, 0.371, 0.824, 1.483 s
CR	4/5
BW	125 kHz
Channels	868.1, 868.3, 868.5, 867.1, 867.3, 867.5, 867.7, 867.9 MHz
Tx Power	14 dBm
Sim. Time	6000 s
Tx Interval	300 s
N Devices	1000
SF allocation	Distance-based
Start time	Between 0 and Tx Interval
Propagation loss	Log-Distance Model
PL_0	7.7 dB
d_0	1 m
γ	3.76
X_g	0

5.2.3 Simulation Results

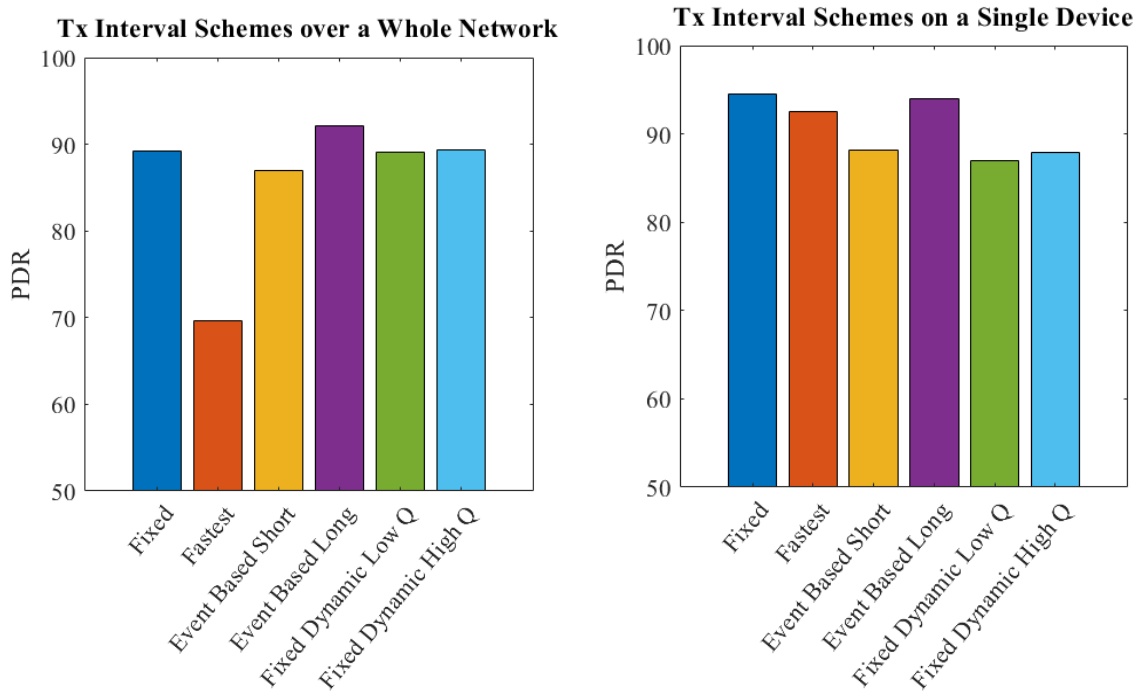
First, network-wide simulations are executed. These assume full control of the network of 1000 devices while applying the same configuration to each node. It is assumed no other devices would be connected to the proprietary network once the simulation is started and the effect of external interference is neglected. While in a real-world urban deployment there is no control over these factors, this is a safe assumption for more rural use cases. Then, the analysis is repeated, focusing on a single device, placed at a distance of 1000 m from the gateway, which parameters are altered exclusively, while the rest of the network operates as described in Section 5.2.2. Each configuration is run 10 times and the obtained PDR results are averaged over the 10 runs, with numerical results reported in Tables 5.6, 5.7 and 5.8 and Fig. 5.6, 5.7 and 5.8.

5.2.4 Varying the Transmission Interval

The first experiment was to vary the transmission schedule of the node by altering the fixed period transmission interval. Five alternative behaviour to the default, which is represented by the "Fixed" case, were coded. The default involves the Tx Interval being set to 300 seconds for each node in the network. It is straight-forward that increasing the time between transmissions reduces the chance of time overlap between packets. Two 1-second packets transmitted from different devices at random points throughout a year have a much lesser chance of colliding than two 1-second packets delivered in a 10-second window. For the simulations concerning a single node the same five behaviours are applied to the lone device, while the remaining 999 behave as per the "Fixed" case.

Below is a rundown of the six behaviours that devices could adopt in this experiment to vary their Tx Interval:

- Fixed: The Tx Interval for all transmissions is set to 300 seconds.
- Fastest: The Tx Interval for all transmissions is set to the fastest possible achievable without going over the duty cycle limit, given the ToA of the previous transmission and the required $T_{silence}$ as explained in Section 2.4.1. For instance, a node set to transmit with SF7 will transmit every 6.11 seconds instead of 300. seconds.
- Event Based Short: The Tx Interval of each transmission is set to be between the fastest the node can transmit and twice that value.
- Event Based Long: The Tx Interval of each transmission is set to be between the Fixed interval (300 seconds) and twice that value (600 seconds).
- Fixed Dynamic Low Q: The Tx Interval of each transmission is equal to the Fixed interval plus or minus a tenth of itself.



(a) Whole network (b) Single device

Figure 5.6: Results from the different Tx Intervals schemes

Table 5.6: Results from 10 simulations per scheme, including the mean PDR, the STD and the output of a T-Test performed between the Fixed case and each other scheme.

	Network-wide					Single Device				
Scheme	Mean	STD	T-Test with Fixed case			Mean	STD	T-Test with Fixed case		
			h	p	t (df)			h	p	t (df)
Fixed	89.23%	0.99	/	/	/	94.5%	15.71	/	/	/
Fastest	69.61%	0.75	1	5.9685e-12	45.5140 (9)	92.5%	7.9	0	0.7527	0.32 (9)
Event Based Short	86.95%	0.53	1	1.0473e-04	6.55 (9)	88.17%	12.94	0	0.4153	0.84 (9)
Event Based Long	92.10%	0.59	1	4.7869e-05	-7.25 (9)	93.94%	8.54	0	0.9306	0.08 (9)
Fixed Dynamic Low Q	89.03%	0.94	0	0.6261	0.50 (9)	87%	17.02	0	0.2443	1.24 (9)
Fixed Dynamic High Q	89.34%	0.3	0	0.7519	-0.32 (9)	87.84%	14.80	0	0.3297	1.03 (9)

- Fixed Dynamic High Q: The Tx Interval of each transmission is equal to the Fixed interval plus or minus 1/30, 1/20, 1/10 or 1/5 of itself.

The "Fixed" case represents standard, periodic operation. The "Fastest" case is set to represent the maximum data transmission rate a device can obtain and meant to model the upper limit for throughput for a device. More realistic operation is that of the two "Event Based" cases. These represent devices which send data only when triggered, for example by a button press or by a light switch, thus exhibiting a non-periodic schedule. Finally, the two "Fixed Dynamic" schemes model a behaviour between periodic and event based, by effectively having a periodic transmission affected by a random time advance or delay.

In the whole network scenario, the only appreciable improvement over operating in the

"Fixed" case is an increase in PDR of 2.87%, obtained by slightly increasing the Tx Interval and randomly picking a value between 300 seconds and 600 seconds ("Event Based Long"). At the same time, reducing the Tx Interval in both the "Fastest" and "Event Based Short" cases lowers the overall PDR, as expected.

Despite the overall lower PDR, these methods can technically still increase the overall amount of data transmitted, since more uplink packets are sent in the same time frame. To keep things fair, in this work each node is restricted to a fixed number of 20 uplinks.

Operating on the whole network, results are found to be statistically different for all schemes except for both of the "Fixed Dynamic" cases. This is understandable given the relative smaller change in behaviour for these two cases when compared to the fixed case. While, for instance, the "Fastest" case can modify the Tx Interval from 300 seconds down to roughly 6.11 seconds for SF7, the "Fixed Dynamic" cases can only ever vary it by a maximum of 60 seconds ($300 \pm 1/5$).

Acting on a single device, as opposed as to every node in the network, no alternative behaviour can yield a benefit over just letting the node transmit with a fixed time of 300 seconds. This is also evident from the results of the t-test, with no scheme operating on the lone device having results which are statistically different from the"Fixed" case. It is also worth noting that 10 samples are usually not enough to provide as accurate a result from the t-test. It is clear from the high standard deviation of the schemes applied to a single node, here and for the experiments presented over the next two sections, that a higher sample size would yield better results.

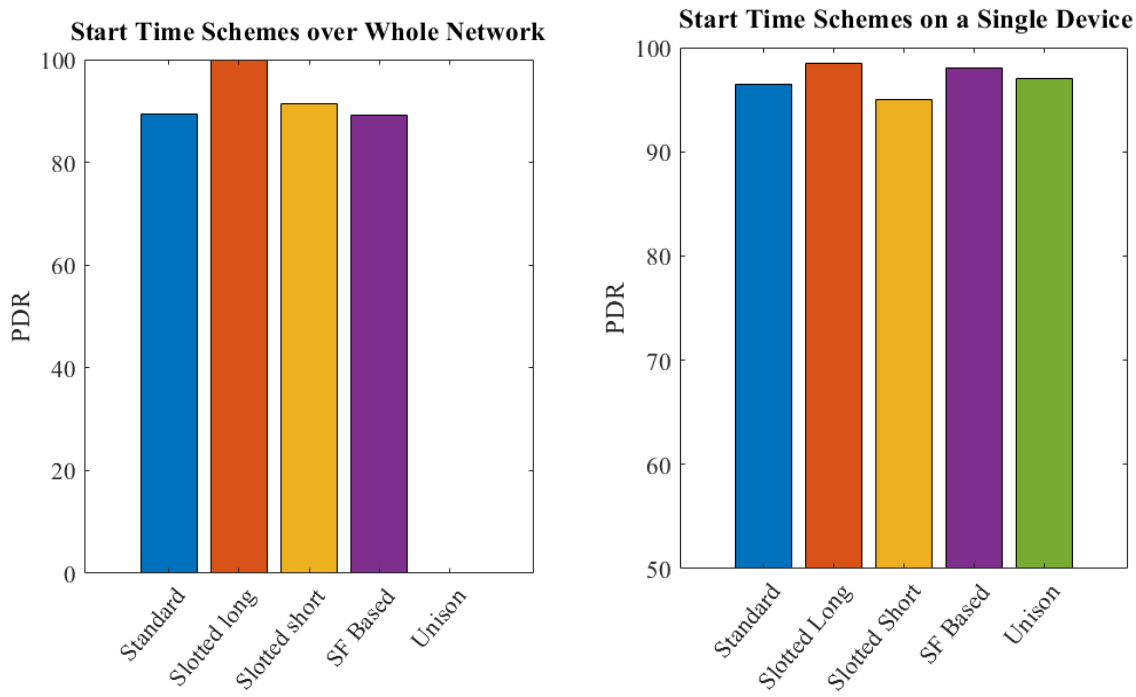
5.2.5 Varying the Device Start Time

Postulating an unrealistic scenario with no external interference and where only a finite, known amount of devices could connect to a network, all operating with periodic Tx Intervals, the total chance of collision could be reduced to 0 by making sure that

$$Tx_{interval} > ToA_{MaxSF} \times \ number\ of\ devices\ seconds \tag{5.4}$$

and by assigning a start time to each device that is at least $ToA_{MaxSF} \times number\ of\ devices$ apart from any other, where ToA_{MaxSF} is the ToA of a packet using the maximum SF in the network under observation. This would in fact prevent packets from ever overlapping in time, which is a condition necessary for packets to collide. Given this, it will be shown in the next experiment that, even when it is not possible to increase the Tx Interval to such extent (our standard network would require a Tx Interval of roughly 1400 seconds), it can still be beneficial for a network of nodes to distance their first (and all the following) transmissions as such.

The devices start time was then varied using the following five methods:



(a) Whole network (b) Single device

Figure 5.7: Results from the different start time schemes

- Standard: each device is set to start transmitting at a random time between 0 and Tx Interval. This ensures that they will all perform the same amount of uplink transmissions during the simulation time, and it models somebody plugging the devices at any time within a single cycle.

- Slotted Long: each node is assigned a start time that depends on its ID and is at a distance in time greater than $ToA_{MaxSF} \times number\ of\ devices$ from the next device starting. For example, if $ToA_{MaxSF} \times number\ of\ devices = 100$ seconds, the node with ID 0 would start transmitting after 0 seconds, the node with ID 1 after 100 seconds, Node 2 after 200 seconds and so on. It is important to note that the Tx Interval of each device in this case is raised to satisfy the inequality at (5.4).

- Slotted Short: each node is assigned a start time that depends on its ID and is at a distance in time greater than $ToA_{MaxSF} \times number\ of\ devices$ from the next transmission. The Tx Interval is not increased, and instead left at the standard 300 seconds.

- SF Based: nodes are set to start transmitting at random times which are the same for each different spreading factor.

- Unison: each node starts transmitting after 1 second.

Table 5.7: Results from 10 simulations per scheme, including the mean PDR, the STD and the output of a T-Test performed between the Standard case and each other scheme.

	Network-wide					Single Device				
Scheme	Mean	STD	T-Test with Standard case			Mean	STD	T-Test with Standard case		
			h	p	t (df)			h	p	t (df)
Standard	89.29%	0.69	/	/	/	96.5%	7.47	/	/	/
Slotted Long	100%	0	1	3.3244e-12	-48.58 (9)	98.5%	3.37	0	0.4945	-0.71 (9)
Slotted Short	91.41%	0.57	1	3.3464e-05	-7.59 (9)	95%	12.69	0	0.7517	0.32 (9)
SF Based	89.15%	0.62	0	0.4969	0.70 (9)	98%	4.83%	0	0.6376	-0.48 (9)
Unison	0%	0	1	1.7299e-20	405.17 (9)	97%	6.32%	0	0.8321	-0.21 (9)

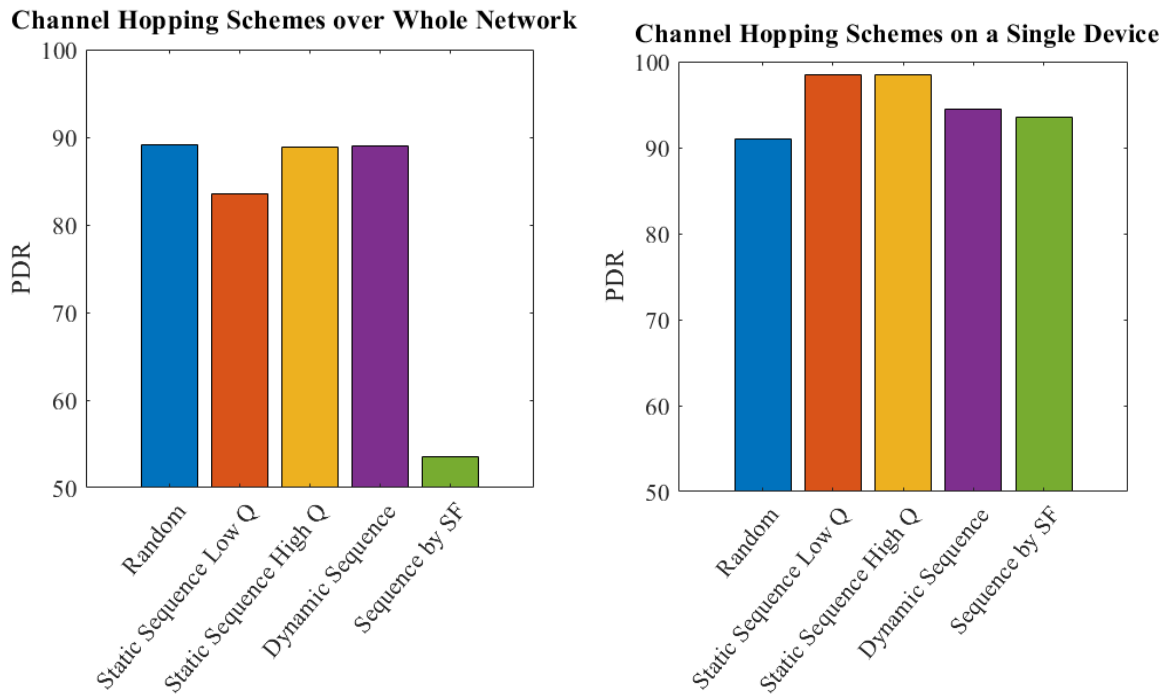
The "Standard" case represent normal operation, with devices being turned on sequentially at random times. The two "Slotted" cases are representative of what can be achieved by leveraging an unique identifier of each device and could be easily applied to any future network. The "SF Based" and "Unison" cases represent theoretically worse operation and are used here for validation.

In the network scenario, "Slotted Long" performed the best, achieving the theoretical 100% PDR, as expected given the relationship at (5.4). More interestingly, slotting start times in the same way while keeping the Tx Interval to 300 seconds in the 'Slotted Short' case, a marginal improvement of 2.12% in PDR is obtained over the performance of the "Standard" assignment. Despite the marginal improvement, these results all pass the t-test, meaning that they can be confidently believed to provide an improvement. The only scheme which is not statistically different from the "Standard" case is the "SF Based" one. This could be due to the small sample size of 10 simulation runs used, despite the result of the mean and standard deviation suggest only a small difference in performance.

Most of the schemes are virtually identical when operating them on a single device, as slotting makes sense only when applied to all devices in a network. In fact, as it is for the case of varying the Tx Interval of a single node here no scheme operating on the lone device yields results which are statistically different from the "Standard" case. Applying a specific start time to a lone device is equivalent to selecting a random start time with different time boundaries. Due to the higher Tx Interval, "Slotted Long" still delivers the greatest PDR, at 98.5%.

5.2.6 Varying the Channel Hopping Behaviour

The final experiment involved testing different methods for a device to decide the next frequency channel to transmit a packet on. In the default configuration this is selected randomly. A check is performed to make sure that transmitting the packet over that channel would not violate the maximum duty cycle limit. If it does not, transmission goes ahead, otherwise a new channel is selected randomly and checked. If all the channels are found to be unavailable to transmit on, the packet transmission is cancelled and not



(a) Whole network (b) Single device

Figure 5.8: Results from the different channel hopping schemes

rescheduled, with the next packet generated being sent instead.

This default method is tested against different sequences of the 8 available channels in the EU frequency plan [57]. Provided the Tx Interval is long enough, each channel can be chosen liberally without restrictions, as it will never violate the duty cycle limit. For these tests, we kept the default Tx Interval of 300 seconds which was chosen appositely for the purpose of staying within the limits.

These methods are:

- Random: devices use the default random hopping between channels described above.
- Static Sequence Low Q: Each device cycles through one of 10 sequences, which are random permutations of the available channels.
- Static Sequence High Q: Each device cycles through one of 1000 sequences, which are random permutations of the available channels.
- Dynamic Sequence: Each device cycles through a permutation of the available 8 channels. This permutation is then changed into another one every time the device reaches the end of it.
- Sequence by SF: Each device cycles through a random permutation of the channels based on its own spreading factor.

Table 5.8: Results from 10 simulations per scheme, including the mean PDR, the STD and the output of a T-Test performed between the Random case and each other scheme.

	Network-wide					Single Device				
Scheme	Mean	STD	T-Test with Random case			Mean	STD	T-Test with Random case		
			h	p	t (df)			h	p	t (df)
Random	89.19%	0.63	/	/	/	91%	10.21	/	/	/
Static Sequence Low Q	83.55%	1.39	1	1.3116e-06	11.27 (9)	98.5%	4.74	1	0.0341	-2.50 (9)
Static Sequence High Q	88.82%	0.78	0	0.3212	1.05 (9)	98.52%	4.74	1	0.0220	-2.76 (9)
Dynamic Sequence	88.99%	0.91	0	0.6483	0.47 (9)	94.5%	8.31	0	0.4823	-0.73 (9)
Sequence by SF	53.49%	2.06	1	7.9645e-13	56.97 (9)	93.5%	14.15	0	0.4854	-0.73 (9)

The "Random" case models the real life behaviour of commercially available equipment. The rest of the schemes are hypothetical improvements based on the idea of distributing and spreading transmissions across all the available channels by leveraging permutations of the available frequencies.

Among the proposed ways to improve performance without relying on downlink, applied network-wide, varying the channel hopping mechanism is the one that yields the least benefits, which is reflected both in the similar mean PDR and the fact that results which are improving on the mean PDR of the "Random" case are not found to be statistically different when performing the t-test. This is because, without knowledge of other devices, transmitting over a random sequence of frequencies is essentially akin to picking frequency channels completely randomly. The hypothesis under test was that, by making sure that every frequency channel would see equal usage via permutations, performance could see an improvement. However, despite the same channel not being able to be selected for transmission multiple times in a row, performance in terms of the network PDR is not improved in any meaningful way.

This is not the case when applying any of our new channel hopping methods to a lone device in a network operating with the default random channel assignment algorithm. In this scenario, any sequence among those tested proves to be more effective than random channel hopping, increasing PDR of the lone device by a maximum of 7.5%. This is confirmed by the fact that the two "Static Sequence" schemes are the only one in this experiment to pass the t-test when applied to a single device. Both a dynamic sequence and a different sequence per SF provide no benefit in terms of PDR and so they are not statistically different.

5.2.7 Summary and Key Findings

In this work a novel way of looking at PDR optimisation was proposed and investigated. No comparable effort was found in LoRaWAN literature at the time of writing, with research being focused on improving the existing downlink reliant methods, instead. The schemes proposed here are downlink-independent and mainly context-agnostic approaches for increasing the PDR of both a network and a single device without relying on commu-

nication with the NS via gateways. Some of the strategies presented in this study only provide a marginal improvement in performance. Longer Tx Intervals giving rise to better PDR is straight forward and confirmed by our results, by boosting the PDR of a 1000 nodes network by roughly 3%. More interestingly, slotting the initial transmission of each device provides a statistically significant improvements of 10.5%. This pushes the original PDR up to a maximum 100% in the theoretical case where each node connecting to a gateway is participating in the scheme. Even assigning a specific channel sequence to a single node in a network is shown to improve its own PDR by roughly 2%. Despite the improvement being marginal, compared to the additional complexity and energy consumption of traditional, downlink-reliant systems, these methods come at a relatively cheap cost of programming the varied behaviour on the network's end devices. The idea of having varied transmission schedules as well as the lessons learned from analysing multi-gateway networks will be key to the work presented in the next chapter. Here, all the modifications and advancement to the NS-3 modules presented in this chapter are put together to create the first recorded instance of a LoRaWAN digital twin of a large, 20-device deployment, reaching 15 different gateways.

Chapter 6

LoRaWAN deployment and its digital twin

The work contained in this chapter is the culmination of the research performed in this PhD and the study of LoRaWAN behaviour in both the theoretical and empirical worlds. As such, it includes all the advancements and modifications to the simulation models discussed in the previous chapters, and it is based on the lessons learned during those studies. It has been published in the IEEE Sensors Journal [26].

6.1 Introduction

Modelling current network installations is a well-established practise in communication, for a variety of technologies and simulators [184, 185]. However, the practise has yet to be broadly applied to LoRa and LoRaWAN, with no previous, LoRaWAN focused work, attempting to create a full digital twin of the same real-life deployment carried out, either working entirely empirically or completely theoretically. As discussed, because of the convenience, large-scale LoRaWAN networks are mostly analysed either through discrete-time simulations or via equivalent analytical models. Despite the initial release of LoRaWAN dating back to 2015, there is still a lack of large-scale performance analysis studies based on real-life deployments or scalable testbeds, with a few exceptions, presented in Section 3.2.2. Even then, most of them consist of only one gateway and a small number of end nodes, with measurements being taken in different locations using a single end device at a time [38, 97, 124]. While real, large-scale networks will always be difficult to install and analyse, the following study seeks, by implementing the digital twin of a LoRaWAN urban deployment, to establish a metric for how successful simulations may be, allowing them to be used more confidently for LoRa and other technologies.

In line with these knowledge gaps, the key contributions of this work are:

- The deployment of a physical 20 nodes network across the University of Glasgow

campus grounds. These nodes operate concurrently, using one of five possible uplink transmission behaviours to simulate a real-world application, with various sensors having varied schedules.

- The porting of the entire physical deployment scenario as close to reality as possible to NS-3, with the creation of a digital twin.
- The analysis of the results of both methods, identifying points of conflict between the two, highlighting further improvement needed to the discrete-time simulation routines and the lessons learnt.

6.2 Network Deployment

6.2.1 System Model

A total of 20 end devices were assembled for this study. Based on the scope of similar research contributions [38] this was deemed to be a sufficient number of concurrently operating devices to make the work relevant, while still being manageable and feasible for a single researcher to build, deploy and manage. Each device is made up of a The Things UNO [186], a modified Arduino Leonardo including a RN2483 chip for LoRaWAN communication, sold by The Things Network [177]. A custom-designed PCB shield is also used to connect a 7.4V, 5.2Ah battery pack to The Things UNO while also providing a quick and handy way of recharging the battery. The device assembly is then placed into weatherproof plastic enclosures, shown in Fig. 6.1. The enclosures were then scattered around the University of Glasgow's 70-acre wide Gilmorehill Campus, in several indoor locations. These varied in altitude, longitude and latitude, each of which was recorded. They were also rooms with different usage, but can be overall regarded as office spaces. None of the nodes were moved throughout the length of the experiment, although external changes in the rooms and buildings were impossible to track. According to previous testing as well as calculations made on the estimated power consumption per uplink and the overall battery capacity, each device was supposed to last on a single charge for approximately 3 to 4 weeks. The GPS and altitude information of each deployed device is available in Table 6.1.

Each node is assigned an ID, ranging from 1 to 20 as well as one of five possible transmission routines, hereby referred to as Tx Groups. These are meant to diversify the deployment uplink schedules and mimic how, in a real-life network deployment, there might be different update needs for different data features to be collected in a single network. For example, a node tasked with recording temperature in a warehouse may need to be monitored once every 10 minutes, while, under normal conditions, a light intensity sensor would only need to transmit data a couple of times during a whole night.

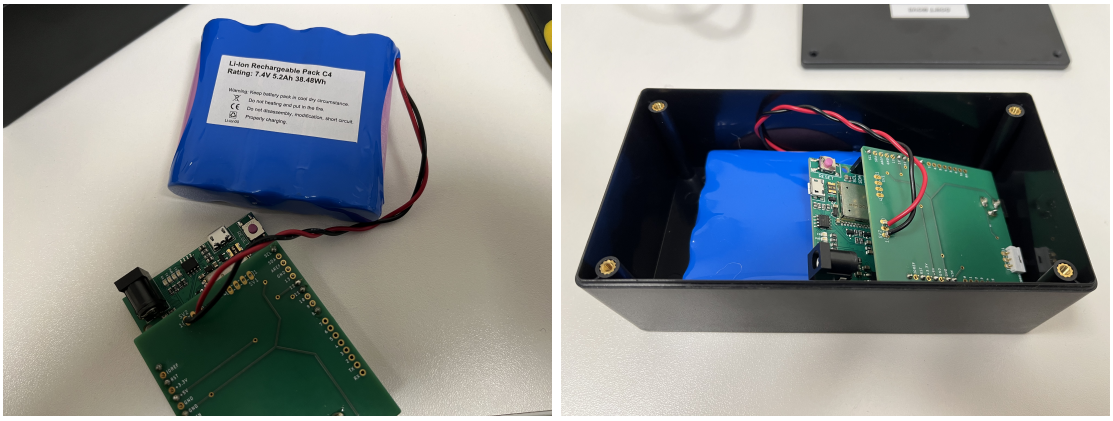


Figure 6.1: One of the nodes deployed and its plastic enclosure

Table 6.1: GPS coordinates and altitude to sea level of the 20 deployed nodes

Node ID	Tx Group	Latitude	Longitude	Altitude (m)
1	1	55.87124255315223	-4.286624829904832	37
2	2	55.87257330986898	-4.285756903792877	32
3	3	55.87243176282027	-4.28585262425438	38
4	4	55.87141134016169	-4.278838648266339	27
5	5	55.872288449209336	-4.286596468130984	32
6	1	55.87237907653764	-4.2865093500802445	33
7	4	55.870815248676585	-4.28703975400883	40
8	3	55.871236462477114	-4.286590852851371	38
9	4	55.87234306825518	-4.286387544053421	33
10	5	55.87101319262039	-4.286845120688884	50
11	1	55.87183100741342	-4.29197002607897	35
12	2	55.872618718379854	-4.293362864890717	30
13	3	55.87165545001698	-4.291994282886463	35
14	4	55.87149441760189	-4.278807799885016	27
15	5	55.87183727538804	-4.291998125418396	36
16	1	55.87141348303532	-4.278834961441753	27
17	2	55.87091417347601	-4.286841088303364	41
18	3	55.87249050849267	-4.2857696050708585	22
19	4	55.8725997	-4.29336376456231	30
20	5	55.87092023604238	-4.286825187775263	41

As far as we are aware, the idea of diversifying the operation of nodes not only based on their data rate, but on their operation and application, has not been considered before. In the available LoRaWAN literature and deployments, unless specified otherwise, nodes have periodic schedules across an entire network with no variation [38,97,119,138]. Additionally, in most of the reviewed deployment, nodes are generally operating singularly instead of concurrently [38, 97]. A breakdown of the potential applications of nodes operating in the different groups is given in Table 6.2.

The Tx groups considered are based on the work performed in Section 5.2. Devices in Tx Group 1 (with ID 1,6,11,16) are programmed to transmit periodically every 300 seconds, those in Tx Group 2 (with ID 2,12,17) are set to transmit as fast as the currently used SF will allow. For devices in Tx Group 3 (ID 3,8,13,18), a random interval between 300 and 900 seconds is chosen for each transmission, and for Tx Group 4 (ID 4,7,9,14,19) a random interval between the fastest the current SF will allow and three times that same value. Finally, devices in Tx Group 5 (ID 5,10,15,20) will cyclically transmit every 600 seconds for 12 hours (representing operation during daytime), then switch to transmitting every 7200 seconds for 12 hours (representing operation during nighttime). These values, which are the ones programmed directly into the nodes via firmware, are also found in Column 4 in Table 6.3.

Due to a number of reasons the actual transmission times recorded upon examining the deployment results were slightly variable, as shown in Column 5 in Table 6.3. One reason for the discrepancy is the inherent hardware delays such as the physical time it takes a device to wake up from sleep and perform the send operation. While a device is in deep sleep, the watchdog timer will check whether to wake it up roughly every 8 seconds, which means that Tx Intervals will be rounded up to the next multiple of 8 seconds. Additionally, the actual send routine takes about 4 seconds to create the packet, set parameters and send the packet. This is why a packet being sent from any device in Tx Group 2 and 4 on SF7 and 8, having calculated and programmed Tx Intervals of 3.1 and 5.64 seconds respectively, will be sent roughly every $8 + 4 = 12$ seconds, instead. As another example, the four devices in Tx Group 1 were found to be transmitting every 304-307, 314-317, 303-307 and 296-299 seconds respectively.

Table 6.2: Nodes in each Tx Group and possible sensing application

Device ID	Tx Group	Possible application
1,6,11,16	1	Standard sensing such as temperature and humidity
2,12,17	2	Monitoring of critical machinery conditions
3,8,13,18	3	Fullness sensing triggered by a pressure plate in the bottom of a bin
4,7,9,14,19	4	Movement sensing triggered by a motion sensor on a barn door
5,10,15,20	5	Solar radiation

Table 6.3: A breakdown of the Tx Interval programmed onto the devices, the Tx Interval obtained from the physical deployment and the Tx Interval obtained from the NS-3 simulation, for different combinations of Tx Groups and SF

Tx Group	SF	ToA (ms)	Programmed Tx Interval per sub-band (s)	Actual Tx Interval (s) from deployment, used as simulation input	Tx Interval (s) obtained at the simulation output
1	7	61.7	300	296 - 317	306 - 307
	8	113.2			
	9	205.8			
	10	370.7			
	11	823.3			
	12	1482.8			
2	7	61.7	3.1	11.95 (avg)	11.99 - 150.14
	8	113.2	5.65	11.95 (avg)	
	9	205.8	10.3	19.08 (avg)	
	10	370.7	18.55	26.95 (avg)	
	11	823.3	41.15	51.96 (avg)	
	12	1482.8	74.15	90.15 (avg)	
3	7	61.7	300 - 900	305 - 911	305 - 915
	8	113.2			
	9	205.8			
	10	370.7			
	11	823.3			
	12	1482.8			
4	7	61.7	3.1-9.3	15.38 (avg)	11.96 - 266.29
	8	113.2	5.64-16.92	20.76 (avg)	
	9	205.8	10.3-30.9	29.50 (avg)	
	10	370.7	18.55-55.65	42.33 (avg)	
	11	823.3	41.15-123.45	95.02 (avg)	
	12	1482.8	74.15-222.45	170 (avg)	
5	7	61.7	500 & 7200	630 & 7462	630 & 7462
	8	113.2			
	9	205.8			
	10	370.7			
	11	823.3			
	12	1482.8			

These averaged values, rather than the ideal ones that were programmed onto the nodes, are then taken as the transmission interval used by the devices when setting up and performing the network simulation, which can be seen to give an output Tx Interval for each combination that is roughly equivalent, in Column 6 in Table 6.3.

Three gateways (ID 1,2,11) were placed in the James Watt South engineering building. Gateways 1 and 2 were placed indoor, on the 4th and 2nd floor, respectively, while gateway 11 was placed on the roof of the same building. It was also expected for the nodes to be deployed to reach a number of third-party gateways in addition to these three gateways. This is given by the fact that the network utilised for this study is managed by IoT Scotland [178] and widely used across the country, with multiple gateways deployed around Glasgow.

Analysis of the results at the end of the deployment will confirm that a total of 15 different gateways received at least one packet from any of our devices. We rely on the accuracy of the GPS position and altitude supplied by the unknown third-party owners of these gateways to put them in space while simulating the network. Location of gateways and nodes are in Figures 6.2 and 6.3[1] as well as Tables 6.1 and 6.4.

Each device starts transmitting using SF12, packets with 10 bytes of payload, plus the 13 bytes of mandatory LoRaWAN overhead. The centre frequency of each transmission is set to hop randomly between 8 possible frequency channels: the mandatory three at 868.1, 868.3, 868.5 MHz and five additional ones at 867.1, 867.3, 867.5, 867.7, 867.9 MHz. The ADR algorithm is activated on both the nodes and the network server, which means

Table 6.4: GPS position and altitude to sea level of all gateways reached during the experiment

GW ID	Latitude	Longitude	Altitude (m)	Location verified
1	55.87119487	-4.287029761	50	Yes
2	55.87091417	-4.286841088	40	Yes
3	55.773037	-4.3326154	140	No
4	55.80429	-4.293961	55	No
5	55.79146	-4.27683	94	No
6	55.860443	-4.2433686	48	No
7	55.84609	-4.3236876	34	No
8	55.873066	-4.2673483	43	No
9	55.82496	-4.2956705	38	No
10	55.80455	-4.2272325	111	No
11	55.870785	-4.2872276	56	Yes
12	55.806576	-4.2837896	61	No
13	55.78999	-4.385697	122	No
14	55.865894	-4.26559	82	No
15	55.781506	-4.266931	95	No

[1]Google Map available here: https://tinyurl.com/2p9npexz

Figure 6.2: Location of all reached gateways and nodes deployed (colour coded by Tx Group) over the entire area

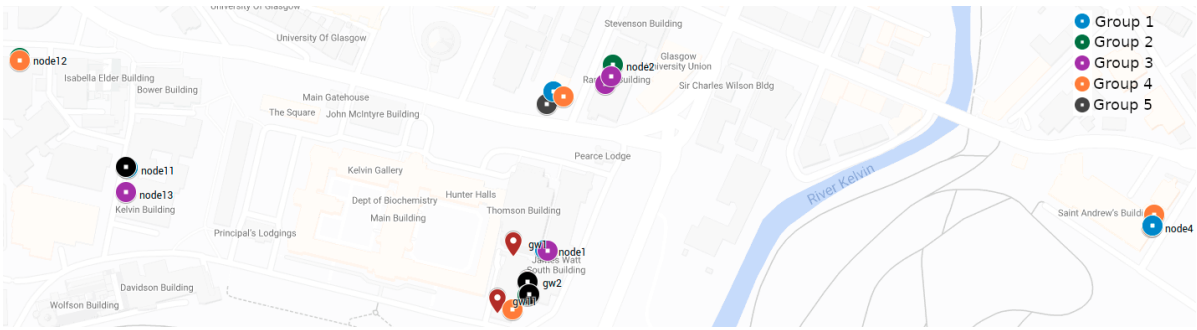


Figure 6.3: A close-up of the University of Glasgow main campus deployment

that the SF and transmitting power of each node may be increased as well as decreased during a node's life cycle. While the on-device ADR algorithm is formally specified and has to be the same for each device, the network-side algorithm is not, with each network operator building their own algorithm based only on suggestions provided directly by Semtech [187]. As far as we are aware, the ADR algorithm used in both the NS-3 digital twin and the actual network server operated by IoT Scotland for this study are based on these recommendations and are equivalent, but because this is proprietary information of the network operator, it cannot be proven and might be the cause of some variance between the deployment and simulation results.

No physical sensors are attached to the devices, partially to extend battery life and partly to describe a generalist network capable of sensing and transmitting any type of data. Instead, the 10 bytes of payload contain information to integrate to the metadata, including *MyCounter*, a custom frame counter keeping track of how many packets were created to be sent, as opposed to the one counting the actually transmitted ones included in the metadata, the ID and Tx Group of the device, and the P_t used for the previous transmission.

The experiment lasted for roughly a full calendar month, with nodes being placed over a period of 48 hours between the 6^{th} and the 8^{th} of December 2021 and picked up during the mid two weeks of January 2022, when it was noticed that no more data was being collected, as all nodes ran out of battery within approximately 3.5 weeks, as expected.

6.2.2 Results

When gathering and analysing the data obtained from an actual LoRaWAN deployment, the only information that can be evaluated is that arriving from correctly received packets, creating the danger of the survivorship bias skewing any conclusions. An important consequence of this is the possible uncertainty regarding a key factor such as the PDR of a network. While one can, to some extent, extrapolate the behaviour of a device based on additional information such as the frame-counter of its last received transmission, there is no guarantee that this packet was the last that was actually transmitted by the node, but only that it was the last one that was correctly received. Consider a node correctly sending its first packet before being placed in a metal enclosure that blocks all further transmissions. Without knowledge of the conditions of the node throughout the deployment and only looking at the results, this node would have a 100% PDR, while in reality many packets were attempted to be transmitted from within the container, but were never received. This is in contrast with simulations, where all packet related events, including their creation and their transmission can be saved and compared to what it is actually received at each gateway and what is then forwarded to the network server.

The devices deployed in the network experienced substantially varying operating times,

as indicated in Table 6.5, even when powered by the same fully charged batteries. These operational periods are calculated using the timestamps of the first and last packets received from each node and are subject to variation for a number of reasons.

Node 14 only delivered 46 packets out of the 158 that its frame counter indicates were attempted during the 7 hours it was definitely active. It most likely failed to keep the connection alive because of an increase in local interference or possibly a misjudged increase in data rate from the network-side ADR that was not afterwards corrected. Other nodes' outage, such as is the case for 20 and 9, which had been working reasonably well during their time alive, may be explained by changes in the building or even room conditions, which could not be continuously monitored. In fact, these discrepancies are mostly a consequence of not having access to up-to-date information on battery status and room conditions and will be included in the lessons learned from this experiment.

To reduce the influence of these unanticipated disparities on the accuracy of a node's PDR, the initial effort was to predict how long each device was meant to be alive for. To this end, the nodes in each Tx Group with the longest time alive were considered to represent the expected duration for that entire group, assuming the same operational time due to the same transmission schedule.

However, as previously stated, Tx Groups 2 and 4 can dynamically change their transmission interval based on their current SF. This feature, along with the fact that the SF

Table 6.5: Nodes ID, groups, operating time, packets sent, delivered and PDR of the deployment

Node ID	Tx Group	Operating hours	Sent	Delivered	PDR (%)
1	1	597	7055	5965	84.55
2	2	42	11997	10313	85.96
3	3	557	3308	2961	89.51
4	4	262	5584	4139	74.12
5	5	511	1573	1227	78
6	1	98	1122	876	78.07
7	4	231	56146	55911	99.58
8	3	533	3178	3150	99.12
9	4	13	2684	2470	92.03
10	5	497	1544	1539	99.68
11	1	589	6958	4259	61.21
12	2	512	20837	6954	33.37
13	3	514	2913	1828	62.75
14	4	7	158	46	29.11
15	5	496	1532	590	38.51
16	1	23	280	266	95
17	2	355	97754	97584	99.83
18	3	529	3191	627	19.65
19	4	321	6827	4749	69.56
20	5	79	248	248	100
TOTAL			234889	205702	87.57

of any device may be adjusted as a result of the ADR algorithm, gives rise to devices in the same Tx Group having vastly different life expectancy because of their different transmission patterns, such as is the case for nodes 12 and 17. This impact is less pronounced in Tx Groups 1, 3, and 5, which have a fixed, SF independent Tx Interval and where the variation in operation time is most likely due to the aforementioned causes, such a misguided ADR command or a change in room condition.

As a result, it was decided to consider the best case scenario instead. Devices are deemed to be active only during the period they were definitely operating, and the simulation is adjusted to make each device stay alive only for that same amount of time. This way the resultant PDR obtained from analysing the physical deployment is unlikely to be an accurate representation of the actual network's Quality-of-Service and throughput. Instead, for the most accurate representation, one should account for the fact that a device should have lasted longer and take all packets that should have been sent during this extra period as lost. Limiting the simulation to the best case scenario, on the other hand, ensures the most fair comparison between simulation and deployment, and as such, it was chosen as the ideal solution to this problem, in this context.

Together with the standard LoRaWAN frame-counter, increasing with every transmission and included in the mandatory packet overhead, we also added a firmware counter, called *MyCounter*, increasing on the device each time a packet was sent to the "send" routine on-board. Discrepancies between *MyCounter* and the regular frame-counter can then be used to check how many packets were created and queued up for transmission but were stopped on the device because of duty cycle restrictions or other reasons.

Taking the *MyCounter* value of the last packet received by a node as the number of packets sent by that node and taking the recorded instances of packets from that device arriving at the network server as the number of packets received will yield a PDR for each node based on:

$$PDR = \frac{Received\ At\ NS}{MyCounter} \times 100\ \% \tag{6.1}$$

The results for each node's PDR calculated as such are shown in Table 6.5.

Upon inspection, it can be seen that on average, the groups with the highest PDR are those with the slowest Tx Interval, as expected. This is not as noticeable here given the fact that "only" 20 devices are being operated concurrently (plus any other nodes outside our test network that we have no knowledge of). The more frequently a node sends, the more likely it is that its packets will collide in the air with others, but as this is an effect that scales with the size of the network it is not as prominent here. Still, we can see that Tx Groups 1 and 5 have roughly 7% higher PDR than Tx Groups 2 and 4, from Fig. 6.4. Tx Group 3, which has average Tx Interval greater than Group 1 as shown in Table 6.3, would perform similarly to Groups 1 and 5. However its average is severely impacted by

node 18, the worst performing of the whole experiment.

Table 6.6 shows the amount of packets received by each gateway from each of the nodes in the deployment. This includes packets that have not yet been deduplicated by the network server and as such the total received from each node is expected to be higher than the number of packets received, as most packets will likely have been received by more than a single gateway. It can be seen that the network as a whole is heavily influenced by mainly 2 devices: gateway 11, receiving the most messages, and the only one receiving at least a packet from every deployed node, and node 17, which similarly, delivers the most messages and is the only node which delivers at least a packet to every gateway. Most importantly, Table 6.6 highlights the impact of gateway and node positioning on their performance. Node 17 and gateway 2 are the pair with the most packets exchanged between them in the uplink direction. They are also the closest, being only 1 metre apart. Node 17 is also the node which packets were received the most by the third-party gateways that our nodes made connection to. The majority of these gateways are to the south of the University Campus and yet node 17 is only the second southernmost device. Node 7 is technically the southernmost node deployed, but it did not enjoy the same high connectivity that devices 17 and 20 did. This is likely because, despite being placed on the same floor and in the same building, nodes 17 and 20 were positioned in a larger,

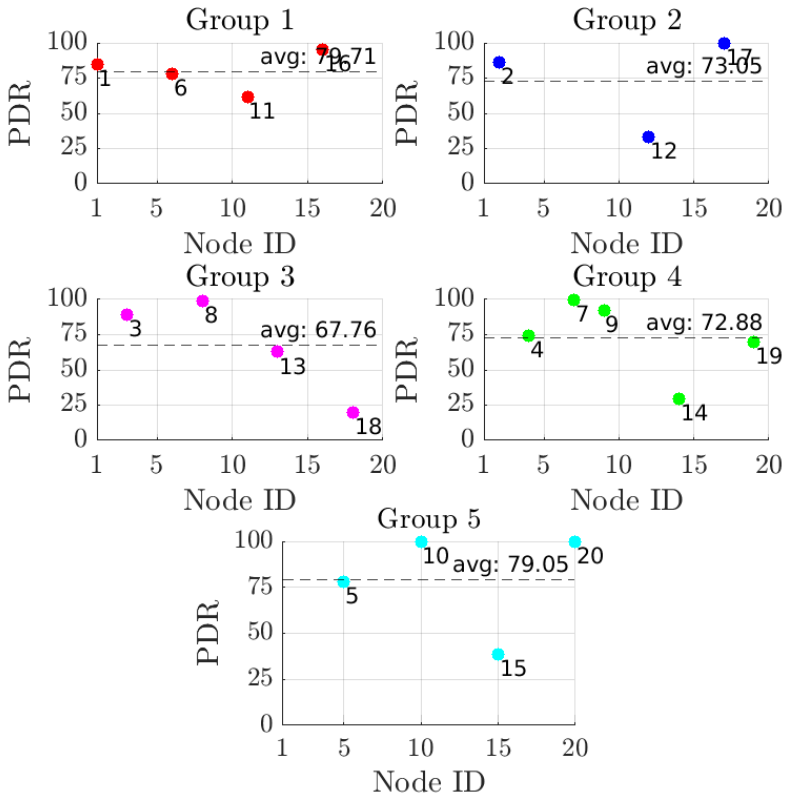


Figure 6.4: PDR of devices per different Tx Groups

Table 6.6: Total number of packets successfully received for each gateway-node pair, including duplicates

	GW 1	GW 2	GW 3	GW 4	GW 5	GW 6	GW 7	GW 8	GW 9	GW 10	GW 11	GW 12	GW 13	GW 14	GW 15
Node 1	341	0	0	0	0	0	0	21	0	0	5920	0	0	0	0
Node 2	475	0	0	0	0	0	0	23	0	0	10271	0	0	13	0
Node 3	831	0	0	0	0	0	3	34	0	0	2848	0	0	62	0
Node 4	0	0	0	0	0	0	0	1789	0	0	3533	0	0	2716	0
Node 5	6	0	0	0	0	0	0	18	0	0	1224	0	0	0	0
Node 6	28	0	0	0	0	0	0	31	0	0	872	0	0	3	0
Node 7	1	54459	0	0	0	10	10	753	0	0	54358	0	0	3	0
Node 8	2457	269	0	0	0	0	0	1869	0	0	2956	0	0	402	0
Node 9	58	1	0	0	0	0	0	19	0	0	2468	0	0	15	0
Node 10	1499	1401	0	0	0	0	0	271	0	0	1472	0	0	93	0
Node 11	0	0	0	0	0	0	0	0	0	0	4259	0	0	11	0
Node 12	0	0	0	0	0	0	0	0	0	0	6785	0	0	1278	0
Node 13	0	0	0	0	0	0	6	0	0	0	1815	0	0	228	0
Node 14	0	0	0	0	0	0	0	41	0	0	1	0	0	13	0
Node 15	0	0	0	0	0	0	0	0	0	0	590	0	0	0	0
Node 16	0	0	0	0	0	0	0	114	0	0	69	0	0	256	0
Node 17	28172	95931	1409	6	206	202	1179	8853	174	1072	95111	3613	20	680	9
Node 18	0	0	0	0	0	0	0	0	0	0	627	0	0	0	0
Node 19	0	0	0	0	0	0	1034	0	0	0	4498	0	3	284	0
Node 20	161	244	172	1	71	9	182	131	58	125	239	216	57	99	0

emptier room with big windows, while node 7 is on a desk in a much smaller, busier and less ventilated room.

In a similar vein, node 14, the one with the shortest time alive, with only 7 active hours recorded, preferred connection to gateways 8 and 14 rather than gateway 11, the most popular one. This is particularly interesting as between node 14 and gateway 11 there is Kelvingrove Park, an open space with no buildings, while between node 14 and gateways 8 and 14 there are sandstone tenements and the expected urban sprawl. This possibly suggests that the conditions of the room and adjacent areas that node 14 was placed in were preventing connection to gateway 11. This is also confirmed by the fact that nodes 4 and 16, placed in the same building as node 14, had a similar experience, although crucially had a better overall connection with gateway 11.

As well as the total amount of packets received, another metric of "usefulness" of each gateway to the network is what we define as their Exclusive Capacity, meaning the amount of packets that were received uniquely by that gateway. While the redundancy obtained by packets being received by multiple gateways is useful and expected, as shown in Table 6.6, the Exclusive Capacity keeps track of packets that, were it not for the gateway that received them, would have been lost. In Fig. 6.6, we can see that gateway 11, the one with the most packets received, also received exclusively 20% of its packets.

Finally, it is always important to keep in mind, as mentioned before, that while the PDR is important, it doesn't always represent the amount of data received. This is clear from Fig. 6.5, where the actual number of packets received per Tx Group is plotted. Devices in Tx Group 2, despite having a lower PDR than those in Tx Groups 1 and 5, have a much faster Tx Interval, so they are responsible for more than 50% of the total data received.

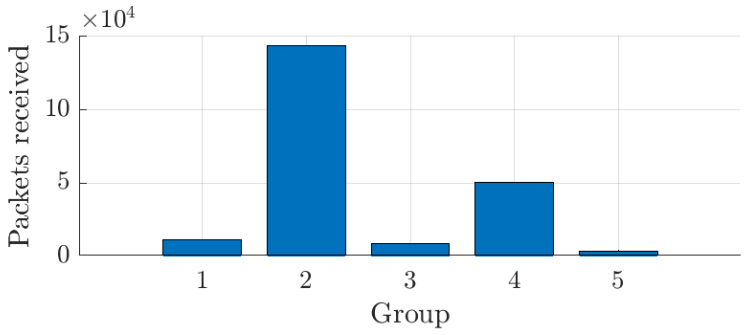


Figure 6.5: Total amount of packets received from nodes in each Tx Group

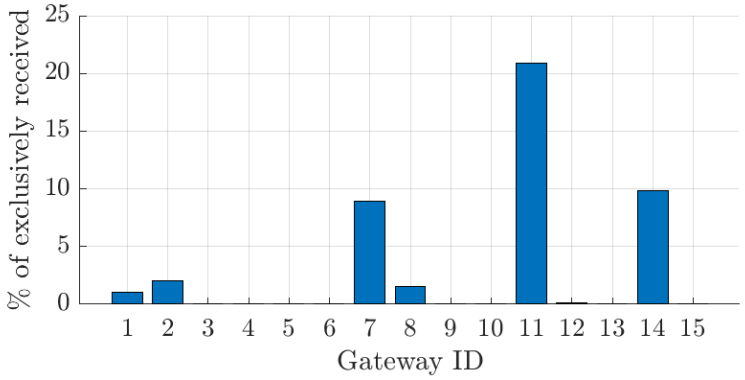


Figure 6.6: Percentage of packets received by each gateway that were exclusively received by that gateway

Deciding which group performance is strictly better is a matter of reliability. In Critical-IoT applications, requiring very strict consistency in terms of the timing of packet arrival, prioritising PDR over total data received is preferable. When not as essential, such as in Massive-IoT applications, transmitting much faster, at the cost of not knowing how reliable the system is, will generally yield a much higher volume of data in the long run and can be preferable. Of course, though energy consumption is not a focus of this work specifically, the choice should also be influenced by the fact that the more data is transmitted, the more energy is used. This is visible in Table 6.5, where devices from Tx Groups 2 and 4 have overall shorter operating times.

6.3 Network Simulation and Comparison

6.3.1 System Model

The deployment is created, modelled and simulated using NS-3, running this thesis author's modified version of the modules first developed in [21] for LoRaWAN analysis. The energy consumption of the devices was not modelled, as outside the scope of this work. Instead, each node is run for the same length of time that it was certainly active during the

network deployment life cycle. This is given by the time difference between the timestamp of its last packet received and its first and is equivalent to the operating time outlined in Table 6.5. ADR is present on both the nodes and the NS in its standard form, as described in the deployment setup.

Nodes and gateways are placed in the 3D space based on the coordinates gathered from positioning the nodes as well as GPS information retrieved from the deployment's results. Approximations for the buildings containing nodes and gateways are also placed in the 3D simulated space. The resulting network topology is visible in Fig. 6.7, and can be visually compared to the actual deployed network as shown by the map in Fig. 6.2.

The core of the simulation is represented by the distance-based path loss in the channel model. For this simulation, choices fell onto the Log-Distance, Okumura-Hata and Friis models. All of these have previously been used to characterise a packet's power loss in LoRaWAN simulations, mathematical models, and empirical data analysis [17, 21, 188–190].

For the Log-Distance case, we used a number of different coefficients for γ and PL_0^{dB}, based on the same formula for urban environments utilised in [21]. Similarly to [191] and [192] we also performed linear regression of the empirically obtained RSSI values from the deployment and obtained γ and PL_0^{dB} coefficients by using the following relation:

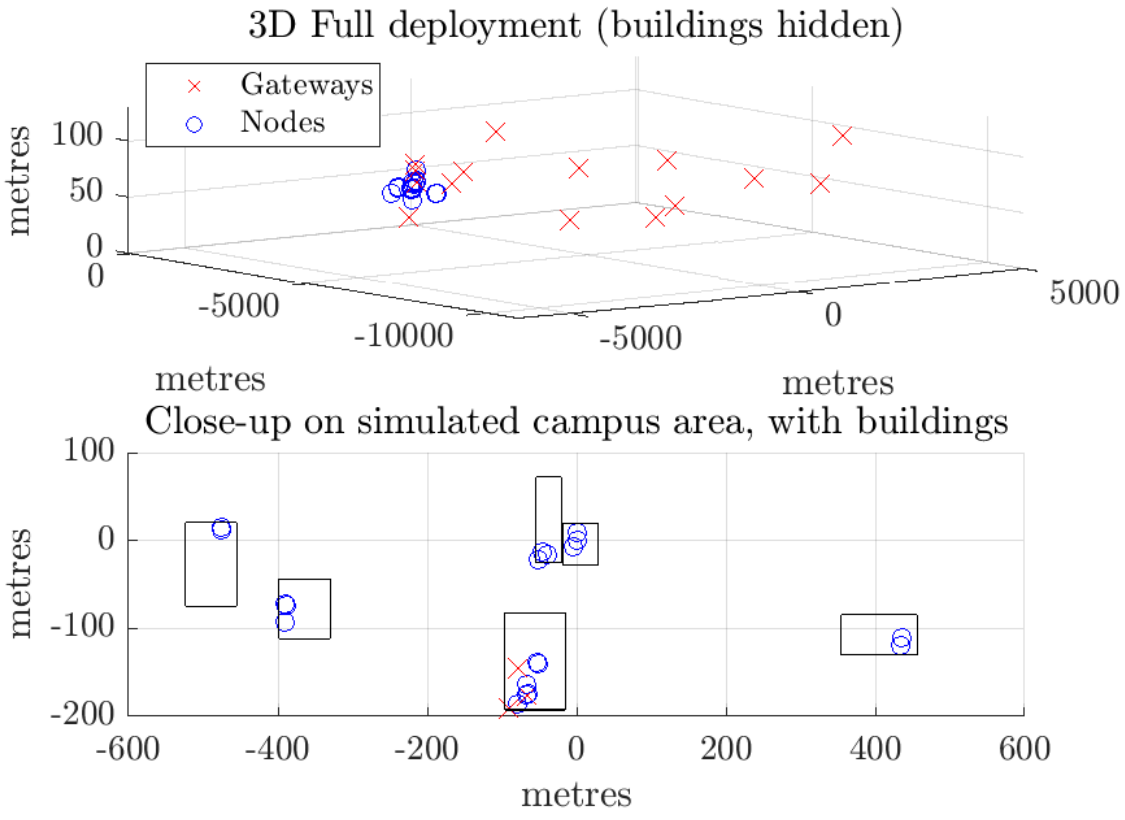


Figure 6.7: The network topology of the deployment as recreated in the NS-3 simulation

$$RSSI = P_t - PL^{dB} + G_{tx} + G_{rx}\ dB \tag{6.2}$$

with

$$PL^{dB} = PL_0^{dB} + 10\gamma log_{10}(d/d_0) + X_g\ dB \tag{6.3}$$

Neglecting the antenna gains G_{tx} and G_{rx} and the fading factor X_g (included by other means in our channel model), and with a standard P_t of 14 dBm and d_0 of 1 metre we get:

$$RSSI = 14 - PL_0^{dB} - 10\gamma log_{10}(d)\ dB \tag{6.4}$$

Having 15 gateways and 20 nodes operating over an area of several kilometres, it made no sense trying to create a unique regression by averaging the data exchanged by each node-gateway pair. Instead, the focus was placed on packets received at gateway 11, as it is the one found to be contributing to the network the most, both in terms of total packets received and packets exclusively received as shown in Fig. 6.6.

The resulting formula for this linear regression, along with the scatter plot of the path loss against the node-gateway distance are shown in Fig. 6.8.

Linked to the main, distance-based propagation loss model, are then included the effects of fast fading, provided by the NS-3 Nakagami-m model, approximating Rayleigh fading, as well as slow fading and building related losses, provided by the original LoRaWAN module [21].

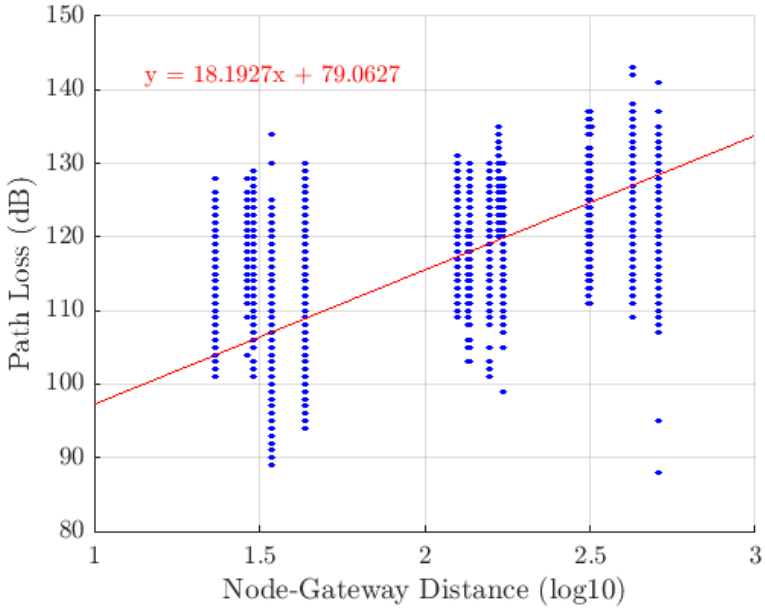


Figure 6.8: Linear fit for the Path Loss recorded at gateway 11

6.3.2 Results and Comparison

Before proceeding with analysing the simulation results and comparing them to the deployment results, the various propagation loss chains were tested to find the most effective and conductive for a proper comparison. This was done by taking the average absolute difference between the PDR of the deployment and the PDR found by performing the simulations using different path loss chains. Table 6.7 shows the results for a selected number of these combinations.

Most of the traditional path loss models such as Log-Distance with default, urban coefficients, as well as Okumura-Hata, have, at best, about a 35% difference in PDR. This makes for a worse digital twin in terms of PDR than what can be obtained by making the simulation perfect, with no packets lost across all devices. This "perfect" configuration in fact yields an average PDR difference of 25.14%.

The two solutions that yield the closest results to the deployed network are a loss chain including the Friis free space model plus the effects of buildings, fast fading and slow fading, and one including the Log-Distance model with the regression coefficients calculated in Fig. 6.8, fast and slow fading, but no buildings effect. The comparison between the PDR of the deployment and that of the simulation running either of these two path loss chains are in Fig. 6.9. This shows the percentage of absolute difference between the two per node, as well as the average and median difference in percentage for both path loss chains.

The Log-Distance regression, unsurprisingly, approximates the deployment best, as we are gathering the best fitting coefficients for γ and PL_0^{dB} (Equation (2.13)) directly from the data that was obtained empirically. Buildings are not taken into account because they

Table 6.7: Some of the PDR differences between simulation and deployment achieved using different Path Loss chains

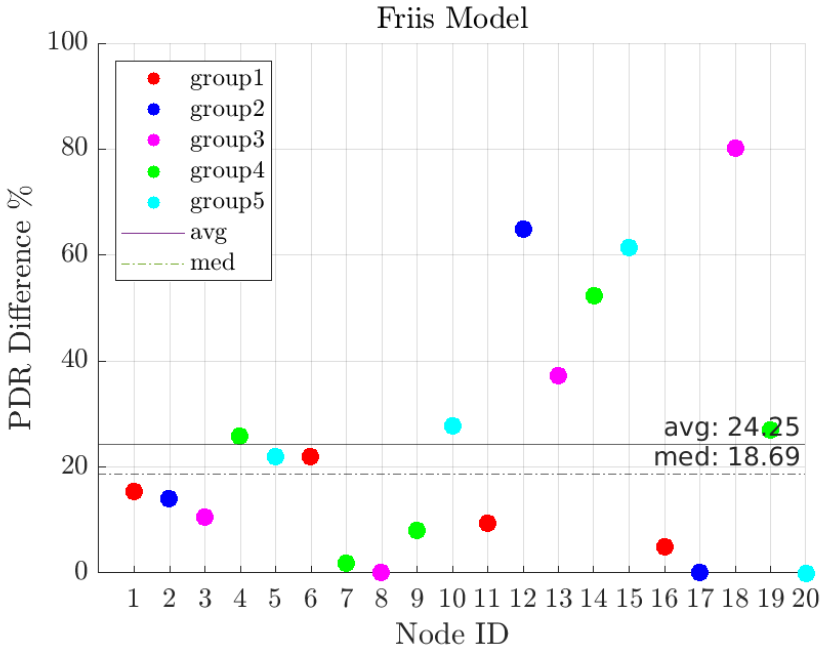
Main propagation loss	Buildings loss	Fast fading	Slow fading	Avg PDR difference (%)
No path loss (perfect network)	X	X	X	25.14
Log-Distance γ =3.76 and $PL_0 = 7.7$ at 1 metre	✓	✓	✓	34.72
Log-Distance γ =3.92 and $PL_0 = 11.52$ at 1 metre	✓	✓	✓	41.15
Log-Distance γ =3.92 and $PL_0 = 11.52$ at 1 metre	✓	✓	X	44.77
Log-Distance (regression) $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	✓	✓	✓	49.46
Log-Distance (regression) $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	X	✓	✓	21.3
Log-Distance (regression) $\gamma = 1.819$ and $PL_0 = 79.063$ at 1 metre	X	X	X	37.13
Okumura-Hata	✓	✓	✓	40.58
Friis	✓	✓	✓	24.25

are already implicitly "included" in the linear regression, while Rayleigh fading as well as correlated slow fading are added to approximate the deviation from that same regressed result. For cases where the regression approach is not feasible, or for more generalised cases, it was found that using a loss chain with Friis, fast and slow fading, as well as building losses, performed the best in this case. From this point on the results of the Log-Distance regression based simulation in Fig. 6.9b will be taken as the default and used to perform the simulation analysis and comparison. A breakdown of the operation and performance of each node in the simulation is given in Table 6.8.

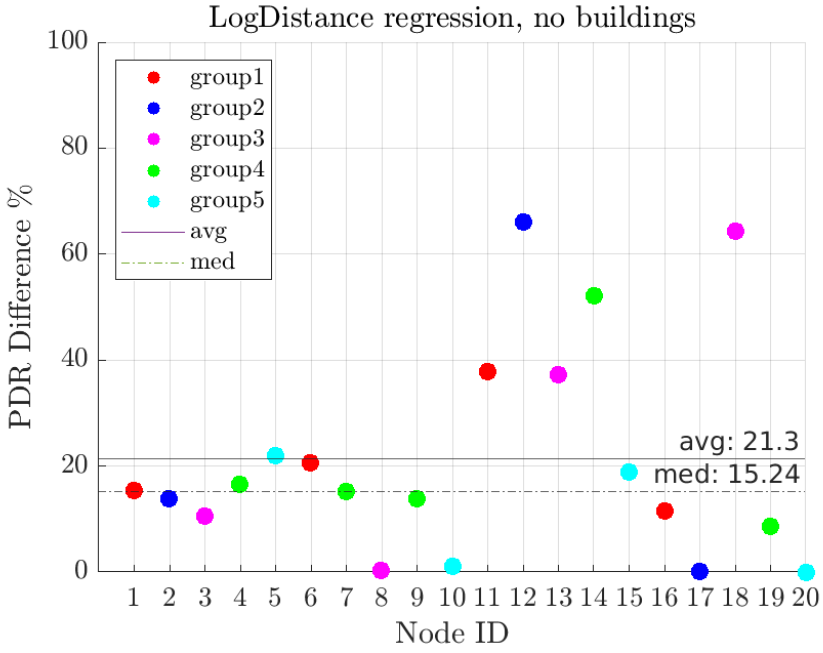
From Fig. 6.9 and Fig. 6.10 emerges that, even when using the best performing path loss chain configurations, some of the nodes have a much worse comparative performance than the others in terms of their PDR. This is the case for nodes 12,14 and 18 all with a PDR difference greater than 50%. As they are not operating in the same Tx Group, this difference in performance between simulation and deployment is too be sought in their positioning, instead. Node 18, for instance, is not only the lowest node deployed in terms of altitude, with an elevation of only 22 m, but also the one in possibly the harshest environment, being placed in a windowless basement room and among electronic manufacturing equipment and machinery. The PDR obtained for it in the actual deployment reflects this, making it the device with the lowest recorded PDR of 19.65% (Table 6.5). Because there is no easy way to inform the simulation about the details of the room in which a node is positioned, node 18 is treated in the same way as others at a comparable distance. As a

Table 6.8: Nodes ID, groups, operating time, packets sent, delivered and PDR of the simulation

Node ID	Tx Group	Operating hours	Sent	Delivered	PDR (%)
1	1	596.95	7047	7043	99.94
2	2	41.99	12571	12534	99.70
3	3	556.79	3309	3308	99.97
4	4	261.99	39194	22547	57.52
5	5	510.59	1574	1574	100
6	1	97.93	1157	1142	98.70
7	4	230.99	34644	29270	84.48
8	3	532.91	3166	3150	99.49
9	4	12.99	1929	1507	78.12
10	5	496.88	1539	1519	98.70
11	1	588.98	6953	6883	98.99
12	2	511.99	153585	152736	99.44
13	3	513.95	3026	3026	100
14	4	6.99	1054	857	81.31
15	5	494.37	1503	863	57.41
16	1	22.95	272	227	83.45
17	2	354.99	106485	106485	100
18	3	528.93	3123	2621	83.92
19	4	320.99	48010	37502	78.11
20	5	78.96	257	257	100
TOTAL			430398	395051	91.78



(a) Friis model, including building effects, fast fading and slow fading



(b) Log-Distance model, using regression coefficients, fast and slow fading and no building effects

Figure 6.9: Average PDR difference between deployment and simulation for the two best propagation loss chains found

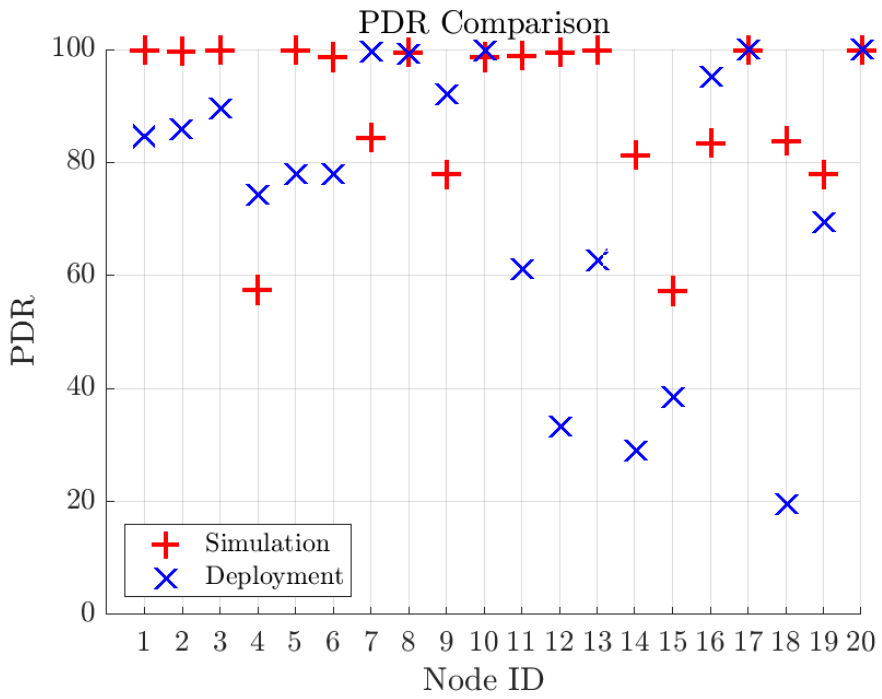


Figure 6.10: Comparison between empirical and simulated PDR

result, its PDR in simulation differs by around 60% between deployment and simulation. Likewise, the opposite is also true, as seen for devices 4, 7, 9, and 16. For these devices, the simulation yields a lower PDR than the real deployment. This is most likely related to the topology of the rooms in which they were put, and it is a natural result of getting the path loss coefficients by linear regression across nodes in different locations.

This is also the case if a path loss chain explicitly including building effects was used, instead. The degree to which buildings may be modelled using the current NS-3 implementation is to provide them with a set of coordinates, an estimated height, and an estimated number of rooms. The simulation is uninformed of the actual room layout, content, and precise position of each node in the room unless significant work was to be carried out to enhance the existing NS-3 models and accurately describe each building involved. Furthermore, the existing building models investigated for this study are not concerned with the actual travel path of a signal, but only with determining whether the receiver and transmitter are indoors, outdoors, in the same building, or in different buildings, to apply different losses regardless of the signal path. Despite these limitations, the impact of the building losses generated by the standard models is still good enough, and in fact, needed for standard, general path loss models. However, adding it to the regression makes this path loss chain the worst performing one, as shown in Table 6.7, with an average PDR difference of 49.46%.

An additional metric to understand how well the simulation is approximating the deployment is obtained by comparing the performance of each gateway in terms of the

Table 6.9: Performance comparison of the gateways in the network

GW ID	Average distance to nodes (metres)	Received in deployment	Received in simulation	Sim/deployment
1	221.51	34029	339992	9.99
2	228.61	152305	307884	2.02
3	11349.23	1581	14	0.009
4	7522.20	7	46	6.57
5	8956.87	277	23	0.083
6	2993.11	221	69	0.31
7	3673.70	2414	69	0.029
8	1227.54	13967	5680	0.41
9	5241.75	232	70	0.30
10	8354.23	1197	26	0.021
11	242.10	199916	292217	1.46
12	7257.50	3829	46	0.012
13	10994.79	80	33	0.4125
14	1484.35	6156	1259	0.204
15	10118.06	9	30	3.33

amount of packets received, which is summarised in Table 6.9.

As it can be seen in Fig. 6.11, the simulation overestimates the packets received at the three closest gateways with IDs 1, 2 and 11. It also underestimates those received at all the others. This is understandable by looking at Fig. 6.8, where most of the residuals from the regression are above the fit over short distances, meaning in reality the deployment suffered a higher Path Loss than the regression suggests, and shift to be below the fit when

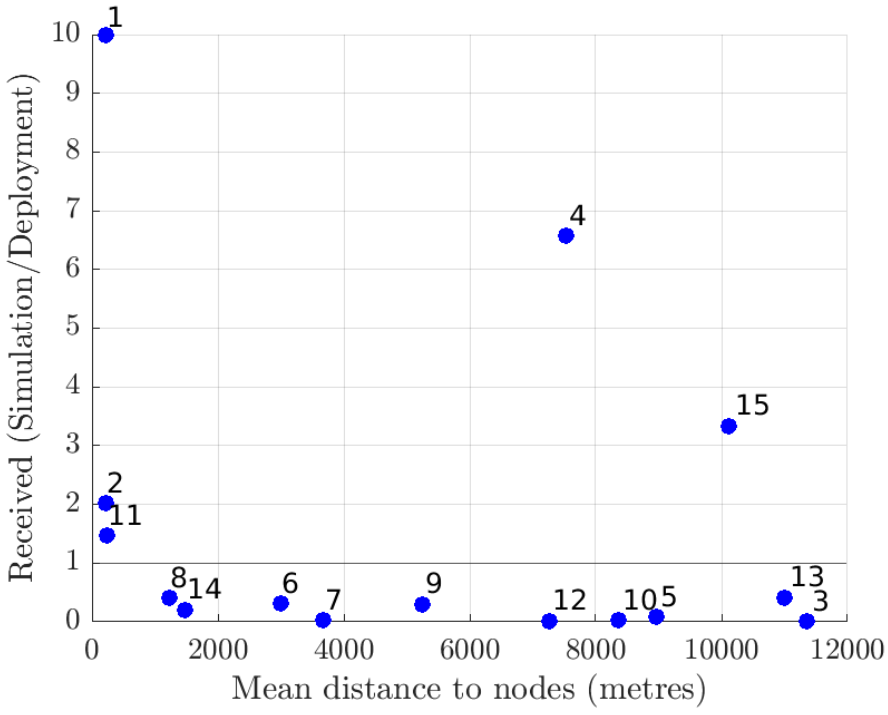


Figure 6.11: Comparison between empirical and simulated gateway performance. Results over 1 indicate that more packets were received for that gateway in simulation than in the deployment

reaching longer link distances. The exception to this, namely gateways 4 and 15, can be explained by looking at the timing of their received packets in Fig. 6.12. This plot shows how these two gateways in particular were probably only active for a smaller amount of time over the course of the one-month experiment.

A fundamental assumption made so far, and applied to the digital twin simulation, is the fact that all gateways stayed active and operating throughout the whole length of the experiment. In reality, other than speculation based on the timing of the reception of packets, as shown in Fig. 6.12, this is not knowable, as only three of those 15 gateways were managed by us. In this same figure, we can see, for instance, that initially all gateways but 4 and 15 were performing well and receiving a good amount of packets. This is while all nodes were still transmitting using SF12, which was the default SF they started with. It is likely that gateway 15 was not on at the start of the deployment as it never received messages from node 20, which died just before gateway 15 received its first packet (black vertical line). Gateway 15 then would go on picking up 9 packets from node 17. The sudden stop of most gateways around the 10^{th} of December is due to the fact that node 17, which as seen in Table 6.6 is the node contributing the most to the deployment, had its data rate switched up to use SF7 by the ADR (red vertical line). What this change does is preventing the farther gateways, which as shown in the same Table 6.6 were only receiving packets from node 17, from receiving any more packets for the rest of the experiment.

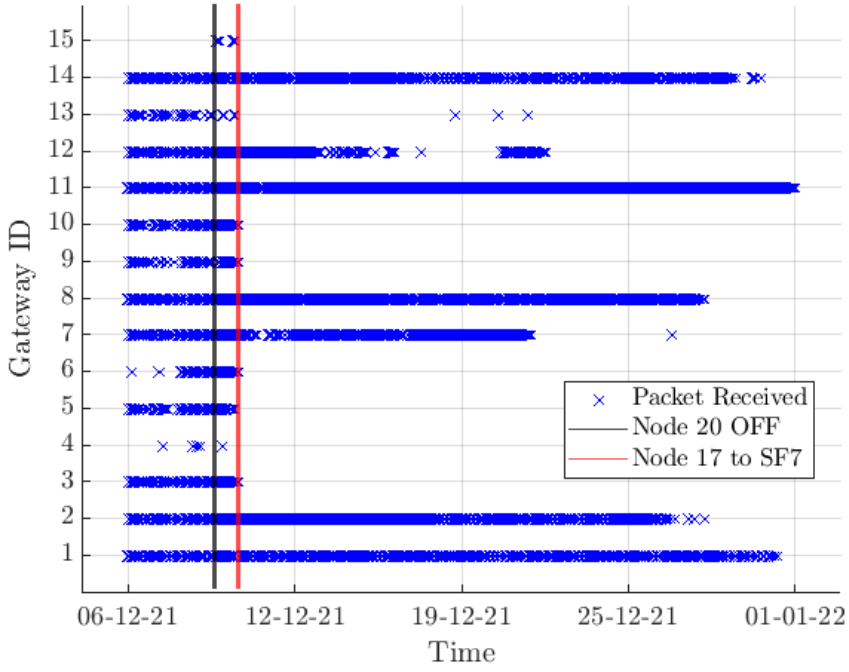


Figure 6.12: Operating time of the 15 gateways detected during the deployment life cycle

6.4 Summary and Key Findings

As part of this study, which is the summation of experience gained over the course of the PhD, a 20 devices LoRaWAN network was deployed in an urban environment and then simulated in software by creating a digital twin to appreciate the differences between the two. The 20 devices are also split into 5 different transmission behaviour groups, to further diversify and characterise the nodes. To the best of our knowledge LoRaWAN digital twins are not yet been investigated in literature.

A few issues arose during this experiment, which would need to be corrected in further work. These are:

- The ADR, while strictly necessary for larger, theoretical networks, in this case caused more issues than it solved. While in the deployment some of the nodes maintained a higher SF and transmission power, in the simulation the ADR seems to lower the SF and transmission power across the board. This is possibly due to the "black box" approach of the ADR implementation on the network server, where no formal requirement is in place. It also created complications with the dynamic scheduling of Tx Groups 2 and 4, as well as limiting the frequency channels to the three mandatory ones after a shift in duty cycle.

- As mentioned, some of the discrepancies also come from the fact that we were unable to determine the actual life cycle of the nodes deployed. While this was not meant to be an energy consumption focused work, in hindsight, the analysis would have greatly benefitted from a way to keep track of the level of charge of each node's battery, so to better predict the expected operating times and actual PDR of the devices. For the future, aside from possibly implementing or using the available energy consumption models for LoRaWAN, at the very least devices should be sending their battery status along with other information in their payload so to make better predictions on their total life span.

- Despite our best efforts, a more controlled environment should be used. Ideally, this would include perfect information on all devices operating in the network and its proximity, their active periods and their exact positioning. In a normal scenario, setting up a private LoRaWAN network rather than piggybacking on one of the open-source ones would allow for more control on the amount of nodes and gateways that are operating at any one time.

- Finally, a building loss model which takes into account the travel path of each signal, computing the approximate number of walls it must travel through, could also benefit this comparison.

Traditional path loss models, even when combined with typical models for additional losses such as building effect, and fast and slow fading are only partially effective in replicating the losses suffered by the urban network we deployed. This difference in PDR can be reduced to around 20% by using empirically derived RSSI data from the deployment itself.

Despite this, we believe the work reported in this chapter is the closest current technology allows simulations to come to a real-world LoRaWAN deployment without using unnecessary resources on exhaustively modelling and replicating every small feature. In reality, even the arguments and lessons learnt raised in this section are only assumed to enhance the comparison, and even if they did, it is difficult to anticipate by how much.

Chapter 7

Conclusions

The research collected and presented in the previous chapters was performed over the past three and a half years. The underlying goal was to analyse and further understand the behaviour of LoRaWAN networks in realistic settings and propose feasible ways to increase LoRaWAN's performance. These enhancements could be potentially applicable to other LPWANs operating similarly to LoRaWAN, especially those operating with an ALOHA like channel access and star-of-stars topology. However, as most of the research surrounding LoRaWAN scalability leverages parameters which are specific to the LoRa modulation technique, such as the spreading factor of a transmission, not all of these ideas will apply.

7.1 Research Contributions

In line with the objectives highlighted in this thesis's introduction, in Section 1.3, the novel contributions to LoRaWAN research that were developed over this course of study can be summarised as follows.

Analysis of LoRaWAN behaviour:
The main activity carried out throughout this research was to increase the general understanding of LoRaWAN behaviour. From being developed as a proprietary technology in 2015, the underlying LoRa modulation has now been successfully reversed-engineered. Its potential and that of MAC Layers such as LoRaWAN, however, is yet to be fully established in the various contexts of IoT. The advancements in this thesis were achieved by studying the available literature, experimenting with physical LoRa hardware and modelling LoRa and LoRaWAN behaviour via software simulations of networks. This effort was the foundation of all other contributions, included, among others, the formalisation of different uplink outage conditions, specifically including duty cycle limits giving rise to "phantom" packets, which are often disregarded but crucial when dealing with commercially available equipment.

Contributing to the enhancement of existing NS-3 models:
This research was also the driver behind a number of enhancements made to existing NS-3 simulation models, which in turn allowed for more realistic LoRaWAN networks to be assessed in simulation, including the different examples reported in this thesis. These enhancements, reported in Section 4.2 have been performed on top of existing open-source modules and made available themselves in open-source. The majority of these advancements were created to either fill gaps in the original ones, or fix incorrect behaviour, and as such they are novel in their usage and implementation in NS-3. The open source nature of the simulator means that they are now freely available to researchers to use and improve even further, thus generating new knowledge.

Comparison between NS-3 and stochastic geometry models:
Both as a validation of the enhanced models and as a method of assessing the similarity between simulation and stochastic geometry when modelling a large-scale network, a novel comparison between the two analysis methods was carried out. Despite their differences, stochastic geometry models and discrete-time simulations are shown to be interchangeable tools for analysing a large-scale network's performance. Furthermore, the changes made to the NS-3 models during the course of this PhD were shown to be functioning as expected by mathematical analysis. New, further model developments, such as the support for more complicated node behaviour, were also inspired by this specific comparison work. Validating the models and proposing a novel comparison with stochastic geometry results will hopefully push more people to investigate relations between the two methods and their linked usage.

Context-unaware, downlink independent optimisation:
The novel idea of context-unaware, downlink independent optimisation was also developed. While most scalability and performance improvements for LoRaWAN rely on downlink communication from the network server to the end-devices, necessarily passing through gateways, in Sec. 5.2 ways of achieving some performance enhancement on each device, while unaware of the network context were tested. Some of the proposed schemes only provide a marginal boost in the measured PDR for nodes and network, though this comes with very little complexity, as opposed to traditional, downlink reliant schemes, which provide a great boost but increasing power consumption and network design complexity to do so. These schemes can be utilised by any LoRaWAN user regardless of network size and resources thus democratising the field and providing an easy and affordable way to get the most out of any node.

LoRaWAN Digital Twin:
As part of this thesis, the first recorded accurate LoRaWAN instance of a digital twin of a deployed network was created and analysed. A network of 20 end-devices was deployed in the urban context of the University Campus and then modelled and simulated in NS-3

by creating a digital twin to be as accurate as possible given the current technology. The 20 nodes were also differentiated by the usage of five different transmission behaviours, modelled according to the work performed in Sec. 5.2. This work was fundamental to highlight areas of necessary improvement in the models moving forward.

7.2 Future Work

Current research in the context of LoRa and LoRaWAN performance and Quality-of-Service can span multiple areas. While this thesis has focused on scalability enhancements, research is also ongoing on issues concerning other aspects of the technology such as security and encryption and power analysis and consumption. Based on the work presented in this thesis, we can highlight the following areas of improvement for future research in a similar context:

Further improvement of the NS-3 modules:
Although enhancements to the originally deployed LoRaWAN modules for NS-3 have been covered in this thesis, there are still areas that are not sufficiently developed. In particular, most calculations performed assume the usage of a fixed Bandwidth of 125kHz, in line with the most commonly used settings for LoRa transmission in Europe. Additionally, Class B and C devices are not fully implemented, with Class A being currently the only end-device class which behaviour can be extensively studied. The support for different ADR and channel access schemes such as TDMA and slotting could also really bolster this area of research. Finally, while the module is specifically designed to recreate LoRaWAN operation, this is not the only MAC layer that can be set up to use LoRa modulation. As the modules feature PHY and MAC behaviour separately, new MAC layers could be introduced and tested by leveraging the existing LoRa model.

Better dynamic parameter allocation:
Whether it is based on channel access modifications such as timeslot scheduling [138, 156] or on ADR-based changes to spreading factors and other transmission parameters [131, 137, 148], downlink reliant schemes are still the best possible solution for commercial, and industrial settings, where network complexity is accepted and power does not come at a premium. The application of machine learning to try to solve optimisation problems featuring multiple gateways and specific scalability-enhancing schemes could lead to the deployment of potentially incrementally better dynamic parameter allocation approaches.

Better context unaware schemes:
Similarly, context-unaware schemes should be further investigated for those applications where little to no control can be gained over the network server, or where power is a very scarce resource. Based on the work presented in Sec. 5.2, future works could expand on this idea by firstly introducing new schemes to alter the transmission parameters dynamically

on each device as well as combining those presented in this thesis. Their application to different scenarios, such as a multi-gateway network, or non-uniform deployments should also be assessed. A mathematical characterisation of each of the methods should also be conducted, including a thorough power consumption analysis, especially when combined with downlink reliant schemes. The different proposed methods could be used concurrently to test the performance of different combinations. Finally, diversity combining [158] and optimal re-transmission algorithms [182] could also be taken into account.

A larger, more controlled Digital Twin:
While the 20 device deployment that was developed and ported as a digital twin in NS-3 was large enough to appreciate some of LoRaWAN characteristics, others were not as prominent, such as packet collision and demodulator's paths saturation. This is due to the relative small amount of concurrently connected devices. A larger network than what is feasible to be built and deployed by a single researcher is something that future work should be aimed at. As mentioned in Section 6.4 for future work one should refrain from using the standard ADR for the sake of a better comparison, as well as deploying a private network to have more control over each device, whether a node or a gateway, that connects to it.

Including energy and security research into the existing analysis:
While out of the scope of this thesis, it would be useful in the future to include into the existing work the effects and constraint arising from other branches of LoRa and LoRaWAN research. In particular, those relating to energy analysis and power consumption [136], as well as encryption and security [193], would add an extra layer of legitimacy to future research. Additionally, this would also provide more reliability to the results, as learned in Section 6.4 in the context of the energy consumption of the nodes deployed in the digital twin.

List of References

[1] S. Agrawal and M. L. Das, "Internet of things - A paradigm shift of future internet applications," *2011 Nirma University International Conference on Engineering: Current Trends in Technology, NUiCONE 2011 - Conference Proceedings*, pp. 8–10, 2011.

[2] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "Iot in smart cities: A survey of technologies, practices and challenges," *Smart Cities*, vol. 4, no. 2, pp. 429–475, 2021.

[3] B. Citoni, F. Fioranelli, M. A. Imran, and Q. H. Abbasi, "Internet of Things and LoRaWAN-Enabled Future Smart Farming," *IEEE Internet of Things Magazine*, vol. 2, no. 4, pp. 14–19, dec 2019.

[4] M. N. Bhuiyan, M. M. Rahman, M. M. Billah, and D. Saha, "Internet of Things (IoT): A Review of Its Enabling Technologies in Healthcare Applications, Standards Protocols, Security, and Market Opportunities," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 474–10 498, 2021.

[5] R. K. Bajaj, M. Rao, and H. Agrawal, "Internet Of Things (IoT) In The Smart Automotive Sector : A Review," *Journal of Computer Engineering*, pp. 36–44, 2018.

[6] P. Ferrari, E. Sisinni, D. F. Carvalho, A. Depari, G. Signoretti, M. Silva, I. Silva, and D. Silva, "On the use of LoRaWAN for the Internet of Intelligent Vehicles in Smart City scenarios," *2020 IEEE Sensors Applications Symposium, SAS 2020 - Proceedings*, pp. 0–5, 2020.

[7] T. R. Wanasinghe, R. G. Gosine, L. A. James, G. K. I. Mann, O. de Silva, and P. J. Warrian, "The internet of things in the oil and gas industry: A systematic review," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8654–8673, 2020.

[8] A. Karmakar, N. Dey, T. Baral, M. Chowdhury, and M. Rehan, "Industrial internet of things: A review," *2019 International Conference on Opto-Electronics and Applied Optics, Optronix 2019*, 2019.

[9] M. Mackenzie, I. Kasujee, and W. Nagy, "IoT forecast: connections, revenue and technology trends 2019–2028," 2018. [Online]. Available: https://www.analysysmason.com/contentassets/f756c575d0d44e9195fc8f7bcd5045bc/analysys_mason_iot_forecast_report_sample_nov2021_rdme0.pdf [Accessed: 30/09/2022]

[10] Fao, "Global agriculture towards 2050," pp. 1–4, 2009. [Online]. Available: http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf [Accessed: 30/09/2022]

[11] The Things Network Global Team, "LoRa World Record Broken: 832km/517mi using 25mW." [Online]. Available: https://www.thethingsnetwork.org/article/lorawan-world-record-broken-twice-in-single-experiment-1 [Accessed: 29/04/2022]

[12] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A study of Lora: Long range & low power networks for the internet of things," *Sensors (Switzerland)*, vol. 16, no. 9, 2016.

[13] R. Sanchez-Iborra and M. D. Cano, "State of the art in LP-WAN solutions for industrial IoT services," *Sensors (Switzerland)*, vol. 16, no. 5, 2016.

[14] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.

[15] LoRa Developer Portal, "What are LoRa® and LoRaWAN®?" [Online]. Available: https://lora-developers.semtech.com/documentation/tech-papers-and-guides/lora-and-lorawan/#:$\sim$:text=Ifeachenddevicesends,additionalgatewaystothenetwork. [Accessed: 25/04/2022]

[16] M. C. Bor, U. Roedig, T. Voigt, and J. 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*, ser. MSWiM '16. New York, NY, USA: ACM, 2016, pp. 59–67.

[17] O. Georgiou and U. Raza, "Low Power Wide Area Network Analysis: Can LoRa Scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, 2017.

[18] R. B. Sørensen, N. Razmi, J. J. Nielsen, and P. Popovski, "Analysis of LoRaWAN Uplink with Multiple Demodulating Paths and Capture Effect," *IEEE International Conference on Communications*, vol. 2019-May, 2019.

[19] A. Mahmood, E. Sisinni, L. Guntupalli, R. Rondon, S. A. Hassan, and M. Gidlund, "Scalability Analysis of a LoRa Network under Imperfect Orthogonality," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1425–1436, 2019.

[20] F. V. Abeele, J. Haxhibeqiri, I. Moerman, and J. Hoebeke, "Scalability analysis of large-scale LoRaWAN networks in ns-3," *arXiv*, vol. 4, no. 6, pp. 2186–2198, 2017.

[21] D. Magrin, M. Centenaro, and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," *IEEE International Conference on Communications*, 2017.

[22] M. Knight and B. Seeber, "Decoding lora: Realizing a modern lpwan with sdr," *Proceedings of the GNU Radio Conference*, vol. 1, no. 1, 2016.

[23] J. Tapparel, "Complete Reverse Engineering of LoRa PHY," *Telecommunications Circuits Laboratory*, 2020.

[24] A. Marquet, N. Montavont, and G. Z. Papadopoulos, "Towards an SDR implementation of LoRa: Reverse-engineering, demodulation strategies and assessment over Rayleigh channel," *Computer Communications*, vol. 153, pp. 595–605, 2020.

[25] B. Citoni, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Impact of Inter-Gateway Distance on LoRaWAN Performance," *Electronics*, vol. 10, no. 18, 2021.

[26] B. Citoni, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Comparative analysis of an urban lorawan deployment: Real world versus simulation," *IEEE Sensors Journal*, vol. 22, no. 17, pp. 17 216–17 223, 2022.

[27] B. Citoni, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Comparative analysis of discrete time simulations and stochastic geometry models of a single gateway lorawan network," in *2021 IEEE International Conference on Smart Internet of Things (SmartIoT)*, 2021, pp. 8–12.

[28] B. Citoni, S. Ansari, Q. H. Abbasi, M. A. Imran, and S. Hussain, "Downlink independent throughput optimisation in lorawan," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022, pp. 1–5.

[29] B. S. Chaudhari, M. Zennaro, and S. Borkar, "LPWAN technologies: Emerging application characteristics, requirements, and design considerations," *Future Internet*, vol. 12, no. 3, 2020.

[30] B. Reynders and S. Pollin, "Chirp spread spectrum as a modulation technique for long range communication," in *2016 Symposium on Communications and Vehicular Technologies (SCVT)*, 2016, pp. 1–5.

[31] H. Rohling and T. May, "Comparison of psk and dpsk modulation in a coded ofdm system," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, vol. 2, 1997, pp. 870–874 vol.2.

[32] S. Gronemeyer and A. McBride, "Msk and offset qpsk modulation," *IEEE Transactions on Communications*, vol. 24, no. 8, pp. 809–820, 1976.

[33] T. Rappaport, "Wireless Communications,Principles and Practice," 2002.

[34] Y. P. Wang, X. Lin, A. Adhikary, A. Grövlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A Primer on 3GPP Narrowband Internet of Things," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117–123, 2017.

[35] B. Foubert and N. Mitton, "Long-range wireless radio technologies: A survey," *Future Internet*, vol. 12, no. 1, 2020.

[36] Design Reuse, "Semtech Acquires Wireless Long Range IP Provider Cycleo." [Online]. Available: https://www.design-reuse.com/news/28706/semtech-cycleo-acquisition.html [Accessed: 2022-09-30]

[37] Y. R. Tsai and J. F. Chang, "Feasibility of combating multipath interference by chirp spread spectrum techniques over Rayleigh and Rician fading channels," *IEEE International Symposium on Spread Spectrum Techniques & Applications*, vol. 1, pp. 282–286, 1994.

[38] J. Haxhibeqiri, E. De Poorter, I. Moerman, and J. Hoebeke, "A survey of LoRaWAN for IoT: From technology to application," *Sensors (Switzerland)*, vol. 18, no. 11, 2018.

[39] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "Impact of LoRa Imperfect Orthogonality: Analysis of Link-Level Performance," *IEEE Communications Letters*, vol. 22, no. 4, pp. 796–799, 2018.

[40] O. Seller and N. Sornin, "European Patent Application EP2763321A1," pp. 1–14, 2014.

[41] LoRa Alliance Technical Commitee, "LoRaWAN 1.1 Specification," p. 101, 2017. [Online]. Available: https://lora-alliance.org/resource-hub/lorawantm-specification-v11 [Accessed: 30/09/2022]

[42] J. S. Sobolewski, "Cyclic redundancy check," in *Encyclopedia of Computer Science*, 2003, pp. 476–479.

[43] J. Haxhibeqiri, F. Van den Abeele, I. Moerman, and J. Hoebeke, "LoRa scalability: A simulation model based on interference measurements," *Sensors (Switzerland)*, vol. 17, no. 6, 2017.

[44] Semtech, "Semtech SX1272." [Online]. Available: https://www.semtech.com/products/wireless-rf/lora-core/sx1272#download-resources [Accessed: 02/05/2022]

[45] Semtech, "Semtech SX1276." [Online]. Available: https://www.semtech.com/products/wireless-rf/lora-core/sx1276 [Accessed: 2022-05-06]

[46] Semtech, "LoRa Modulation Basics AN1200.22," *App Note*, no. May, pp. 1–26, 2015. [Online]. Available: http://www.semtech.com/images/datasheet/an1200.22.pdf [Accessed: 04/05/2022]

[47] M. Cattani, C. A. Boano, and K. Römer, "An Experimental Evaluation of the Reliability of LoRa Long-Range Low-Power Wireless Communication," *Journal of Sensor and Actuator Networks*, vol. 6, no. 2, p. 7, 2017.

[48] Semtech, "SX1272/73 - 860 MHz to 1020 MHz Low Power Long Range Transceiver | Semtech," p. 129, 2019. [Online]. Available: http://www.semtech.com/wireless-rf/rf-transceivers/sx1276/ [Accessed: 02/05/2022]

[49] Semtech, "SX1276/77/78/79 - 137 MHz to 1020 MHz Low Power Long Range Transceiver," 2020. [Online]. Available: https://semtech.my.salesforce.com/sfc/p/#E0000000JelG/a/2R0000001Rbr/6EfVZUorrpoKFfvaF_Fkpgp5kzjiNyiAbqcpqh9qSjE [Accessed: 02/05/2022]

[50] Semtech, "SX1272/3/6/7/8: LoRa Modem, Designer's Guide, AN1200.13," 2013. [Online]. Available: https://www.openhacks.com/uploadsproductos/loradesignguide_std.pdf [Accessed: 30/09/2022]

[51] LoRa Alliance, "LoRaWAN® Regional Parameters 1.1," pp. 1–56, 2017.

[52] J. C. Liando, A. Gamage, A. W. Tengourtius, and M. Li, "Known and unknown facts of LoRa: Experiences from a large-scale measurement study," *ACM Transactions on Sensor Networks*, vol. 15, no. 2, 2019.

[53] F. Benkhelifa, Y. Bouazizi, and J. A. McCann, "How Orthogonal is LoRa Modulation?" *IEEE Internet of Things Journal*, vol. 0, no. 0, pp. 1–1, 2022.

[54] M. Bor, J. Vidler, and U. Roedig, "Lora for the internet of things," in *Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*, ser. EWSN '16. USA: Junction Publishing, 2016, pp. 361–366.

[55] B. Ray, "Improving LoRaWAN Scalability," 2017, https://www.link-labs.com/blog/improving-lorawan-scalability [Online; accessed 31-May-2019].

[56] Semtech, "LoRa and LoRaWAN: A Technical Overview," *Semtech Technique Paper*, no. December 2019, pp. 1–17, 2020. [Online]. Available: https://lora-developers.semtech.com/library/tech-papers-and-guides/lora-and-lorawan/ [Accessed: 30/09/2022]

[57] The Things Network, "LoRaWAN." [Online]. Available: https://www.thethingsnetwork.org/docs/lorawan/ [Accessed: 30/09/2022]

[58] P. S. Cheong, J. Bergs, C. Hawinkel, and J. Famaey, "Comparison of LoRaWAN classes and their power consumption," *2017 IEEE Symposium on Communications and Vehicular Technology, SCVT 2017*, vol. 2017-December, pp. 1–6, 2017.

[59] F. Delobel, N. El Rachkidy, and A. Guitton, "Analysis of the Delay of Confirmed Downlink Frames in Class B of LoRaWAN," *IEEE Vehicular Technology Conference*, vol. 2017-June, no. 1, 2017.

[60] D. Ron, C. J. Lee, K. Lee, H. H. Choi, and J. R. Lee, "Performance Analysis and Optimization of Downlink Transmission in LoRaWAN Class B Mode," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7836–7847, 2020.

[61] L. Polak and J. Milos, "Performance analysis of LoRa in the 2.4 GHz ISM band: coexistence issues with Wi-Fi," *Telecommunication Systems*, vol. 74, no. 3, pp. 299–309, 2020.

[62] Electronic Communications Committee, "ERC Recommendation 70-03," pp. 1–78, 2015. [Online]. Available: https://docdb.cept.org/download/25c41779-cd6e/Rec7003e.pdf [Accessed: 02/05/2022]

[63] Semtech, "AN1200.10: ETSI Compliance of the SX1272/3 LoRa Modem. Rev. 1," pp. 1–25, 2013.

[64] ETSI, "Electromagnetic compatibility and Radio spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment to be used in the 25 MHz to 1 000 MHz frequency range with power levels ranging up to 500 mW;," pp. 1–73, 2014. [Online]. Available: https://www.etsi.org/deliver/etsi_en/300200_300299/30022001/02.04.01_40/en_30022001v020401o.pdf [Accessed: 02/05/2022]

[65] Ofcom, "IR 2030 – UK Interface Requirements 2030," 2021. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0028/84970/ir-2030.pdf [Accessed: 30/09/2022]

[66] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of lorawan," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, Sep. 2017.

[67] J. Finnegan, "Duty cycle regulations in Europe," 2017. [Online]. Available: https://www.thethingsnetwork.org/forum/t/duty-cycle-regulations-in-europe/5015 [Accessed: 02/05/2022]

[68] Microchip, "RN2483," pp. 1–22, 2019.

[69] N. Abramson, "THE ALOHA SYSTEM," in *Proceedings of the November 17-19, 1970, fall joint computer conference on - AFIPS '70 (Fall)*, vol. 25, no. 1. New York, New York, USA: ACM Press, 1970, p. 281.

[70] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, fifth edit ed. Pearson, 2011.

[71] C. Goursaud and J. M. Gorce, "Dedicated networks for IoT: PHY / MAC state of the art and challenges," *EAI Endorsed Transactions on Internet of Things*, vol. 1, no. 1, p. 150597, 2015.

[72] J. M. Marais, A. M. Abu-Mahfouz, and G. P. Hancke, "A Review of LoRaWAN Simulators: Design Requirements and Limitations," *Proceedings - 2019 International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2019*, pp. 1–6, 2019.

[73] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, pp. 89–97, 2018.

[74] Semtech, "Semtech SX1301." [Online]. Available: https://www.semtech.com/products/wireless-rf/lora-core/sx1301 [Accessed: 04/05/2022]

[75] M. Capuzzo, D. Magrin, and A. Zanella, "Mathematical Modeling of LoRa WAN Performance with Bi-directional Traffic," *2018 IEEE Global Communications Conference, GLOBECOM 2018 - Proceedings*, 2018.

[76] Q. L. Hoang, V. N. Q. Bao, and H. Oh, "Performance Evaluation of LoRa Networks for Confirmed Messages," in *2021 International Conference on Advanced Technologies for Communications (ATC)*. IEEE, oct 2021, pp. 162–166.

[77] Semtech, "LoRaWAN – simple rate adaptation recommended algorithm," pp. 1–8, 2016.

[78] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[79] Y. Okumura, "Field strength and its variability in vhf and uhf land-mobile radio service," *Rev. Electr. Commun. Lab.*, vol. 16, pp. 825–873, 1968.

[80] M. Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, 1980.

[81] Semtech, "LoRa Applications." [Online]. Available: https://www.semtech.com/lora/lora-applications [Accessed: 2022-05-05]

[82] M. Usmonov and F. Gregoretti, "Design and implementation of a lora based wireless control for drip irrigation systems," in *2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)*, Dec 2017, pp. 248–253.

[83] B. R. Dos Reis, Z. Easton, R. R. White, and D. Fuka, "A LoRa sensor network for monitoring pastured livestock location and activity," *Translational Animal Science*, vol. 5, no. 2, pp. 1–9, 2021.

[84] R. X. W. Zhao, S. Lin, J. Han and L. Hou, "Design and Implementation of Smart Irrigation System Based on LoRa," *2017 IEEE Globecom*, pp. pp. 1–6, 2017.

[85] J. D. Silveira, A. F. Artur, J. V. Dos Reis Junior, A. C. Soares, and R. A. Rabelo, "A New Low-Cost LoRaWAN Power Switch for Smart Farm Applications," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pp. 3330–3335, 2021.

[86] R. K. Singh, M. Aernouts, M. De Meyer, M. Weyn, and R. Berkvens, "Leveraging LoRaWAN technology for precision agriculture in greenhouses," *Sensors (Switzerland)*, vol. 20, no. 7, 2020.

[87] N. H. N. Ibrahim, A. R. Brahim, I. Mat, A. N. Harun, and G. Witjaksono, "IR 4.0 using IoT and LORAWAN to Accelerate Lentinula Edodes Growth," in *2018 2nd International Conference on Smart Sensors and Application (ICSSA)*. IEEE, jul 2018, pp. 28–32.

[88] J. P. Amaro, H. Marcos, and R. Cortesao, "An Energy Study of a LoRaWan Based Electrical Impedance Spectroscopy Module for Tree Health Monitoring," *IECON Proceedings (Industrial Electronics Conference)*, vol. 2019-Octob, pp. 2916–2921, 2019.

[89] V. Sharma, I. You, G. Pau, M. Collotta, J. D. Lim, and J. N. Kim, "LoRaWAN-based energy-efficient surveillance by drones for intelligent transportation systems," *Energies*, vol. 11, no. 3, 2018.

[90] P. Manzoni, C. T. Calafate, J. C. Cano, and E. Hernández-Orallo, "Indoor vehicles geolocalization using LoRaWAN," *Future Internet*, vol. 11, no. 6, 2019.

[91] Y. Li, J. Barthelemy, S. Sun, P. Perez, and B. Moran, "Urban vehicle localization in public LoRaWan network," *IEEE Internet of Things Journal*, vol. 14, no. 8, pp. 1–1, 2021.

[92] F. Bonafini, D. Fernandes Carvalho, A. Depari, P. Ferrari, A. Flammini, M. Pasetti, S. Rinaldi, and E. Sisinni, "Evaluating indoor and outdoor localization services for LoRaWAN in Smart City applications," *2019 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2019 - Proceedings*, pp. 300–305, 2019.

[93] G. Del Campo, I. Gomez, S. Calatrava, R. Martinez, and A. Santamaria, "Power distribution monitoring using lora: Coverage analysis in suburban areas," *International Conference on Embedded Wireless Systems and Networks*, pp. 233–238, 2018.

[94] H. Klaina, I. P. Guembe, P. Lopez-Iturri, J. J. Astrain, L. Azpilicueta, O. Aghzout, A. V. Alejos, and F. Falcone, "Aggregator to Electric Vehicle LoRaWAN Based Communication Analysis in Vehicle-to-Grid Systems in Smart Cities," *IEEE Access*, vol. 8, pp. 124 688–124 701, 2020.

[95] A. Candia, S. N. Represa, D. Giuliani, M. A. Luengo, A. A. Porta, and L. A. Marrone, "Solutions for SmartCities: Proposal of a monitoring system of air quality based on a LoRaWAN network with low-cost sensors," *Congreso Argentino de Ciencias de la Informatica y Desarrollos de Investigacion, CACIDI 2018*, 2018.

[96] M. Y. Thu, W. Htun, Y. L. Aung, P. E. E. Shwe, and N. M. Tun, "Smart air quality monitoring system with LoRaWAN," *Proceedings - 2018 IEEE International Conference on Internet of Things and Intelligence System, IOTAIS 2018*, pp. 10–15, 2019.

[97] J. Petäjäjärvi, K. Mikhaylov, M. Hämäläinen, and J. Iinatti, "Evaluation of lora lpwan technology for remote health and wellbeing monitoring," in *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*, March 2016, pp. 1–5.

[98] A. Mdhaffar, T. Chaari, K. Larbi, M. Jmaiel, and B. Freisleben, "IoT-based health monitoring via LoRaWAN," *17th IEEE International Conference on Smart Technologies, EUROCON 2017 - Conference Proceedings*, no. July, pp. 519–524, 2017.

[99] C. D. Fernandes, A. Depari, E. Sisinni, P. Ferrari, A. Flammini, S. Rinaldi, and M. Pasetti, "Hybrid indoor and outdoor localization for elderly care applications with LoRaWAN," *IEEE Medical Measurements and Applications, MeMeA 2020 - Conference Proceedings*, pp. 5–10, 2020.

[100] P. A. Catherwood, D. Steele, M. Little, S. McComb, and J. McLaughlin, "A Community-Based IoT Personalized Wireless Healthcare Solution Trial," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, no. March, pp. 1–13, 2018.

[101] N. C. Gaitan, "A long-distance communication architecture for medical devices based on lorawan protocol," *Electronics (Switzerland)*, vol. 10, no. 8, 2021.

[102] J. Haxhibeqiri, A. Karaagac, F. Van Den Abeele, W. Joseph, I. Moerman, and J. Hoebeke, "LoRa indoor coverage and performance in an industrial environment: Case study," in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2017, pp. 1–8.

[103] G. Loubet, A. Takacs, E. Gardner, A. De Luca, F. Udrea, and D. Dragomirescu, "LoRaWAN battery-free wireless sensors network designed for structural health monitoring in the construction domain," *Sensors (Switzerland)*, vol. 19, no. 7, 2019.

[104] R. Dai, O. Diraneyya, and S. Brell-Çokcan, "Improving data communication on construction sites via LoRaWAN," *Construction Robotics*, vol. 5, no. 2, pp. 87–100, 2021.

[105] T. Porselvi, C. S. Sai Ganesh, B. Janaki, K. Priyadarshini, and S. Shajitha Begam, "IoT based coal mine safety and health monitoring system using LoRaWAN," *2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021*, no. May, pp. 49–53, 2021.

[106] I. Bizon Franco de Almeida, M. Chafii, A. Nimr, and G. Fettweis, "Alternative Chirp Spread Spectrum Techniques for LPWANs," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 4, pp. 1846–1855, dec 2021.

[107] R. Bomfin, M. Chafii, and G. Fettweis, "A novel modulation for IoT: PSK-LoRa," *IEEE Vehicular Technology Conference*, vol. 2019-April, 2019.

[108] T. Elshabrawy and J. Robert, "Interleaved chirp spreading LoRa-based modulation," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3855–3863, 2019.

[109] P. Edward, E. Tarek, M. El-Aasser, M. Ashour, and T. Elshabrawy, "Further LoRa Capacity Enhancement through Interleaved Chirp Spreading LoRa Expansion," *International Conference on Wireless and Mobile Computing, Networking and Communications*, vol. 2019-Octob, 2019.

[110] P. Edward, M. El-Aasser, M. Ashour, and T. Elshabrawy, "Interleaved Chirp Spreading LoRa as a Parallel Network to Enhance LoRa Capacity," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3864–3874, 2021.

[111] M. Hanif and H. H. Nguyen, "Slope-Shift Keying LoRa-Based Modulation," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 211–221, jan 2021.

[112] A. Mondal, M. Hanif, and H. H. Nguyen, "SSK-ICS LoRa: A LoRa-Based Modulation Scheme with Constant Envelope and Enhanced Data Rate," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1185–1189, 2022.

[113] M. Noor-A-Rahim, M. O. Khyam, A. Mahmud, X. Li, D. Pesch, and H. V. Poor, "Hybrid Chirp Signal Design for Improved Long-Range (LoRa) Communications," *Signals*, vol. 3, no. 1, pp. 1–10, 2022.

[114] P. Edward, A. Muhammad, S. Elzeiny, M. Ashour, T. Elshabrawy, and J. Robert, "Enhancing the Capture Capabilities of LoRa Receivers," *2019 International Conference on Smart Applications, Communications and Networking, SmartNets 2019*, pp. 2–7, 2019.

[115] Lukas, W. A. Tanumihardja, and E. Gunawan, "On the application of IoT: Monitoring of troughs water level using WSN," *2015 IEEE Conference on Wireless Sensors, ICWiSE 2015*, pp. 58–62, 2016.

[116] S. Benaissa, D. Plets, E. Tanghe, J. Trogh, L. Martens, L. Vandaele, L. Verloock, F. A. Tuyttens, B. Sonck, and W. Joseph, "Internet of animals: Characterisation of LoRa sub-GHz off-body wireless channel in dairy barns," *Electronics Letters*, vol. 53, no. 18, pp. 1281–1283, 2017.

[117] A. Grunwald, M. Schaarschmidt, and C. Westerkamp, "LoRaWAN in a rural context: Use cases and opportunities for agricultural businesses," in *24. ITG-Symposium on Mobile Communication - Technologies and Applications*, 2020, pp. 134–139.

[118] D. Yim, J. Chung, Y. Cho, H. Song, D. Jin, S. Kim, S. Ko, A. Smith, and A. Riegsecker, "An Experimental LoRa Performance Evaluation in Tree Farm," *2018 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2018.

[119] R. Yasmin, J. Petajajarvi, K. Mikhaylov, and A. Pouttu, "Large and Dense LoRaWAN Deployment to Monitor Real Estate Conditions and Utilization Rate," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2018-Septe, 2018.

[120] R. Muppala, A. Navnit, S. Poondla, and A. M. Hussain, "Investigation of Indoor LoRaWAN Signal Propagation for Real-World Applications," *2021 6th International Conference for Convergence in Technology, I2CT 2021*, pp. 17–21, 2021.

[121] M. Ohta, K. Adachi, N. Aihara, O. Takyu, T. Fujii, and M. Taromaru, "Measurement Experiments on 920 MHz Band for Spectrum Sharing with LoRaWAN," *IEEE Vehicular Technology Conference*, vol. 2018-Augus, 2018.

[122] A. M. Yousuf, E. M. Rochester, and M. Ghaderi, "A low-cost lorawan testbed for iot: Implementation and measurements," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Feb 2018, pp. 361–366.

[123] Q. Liu, Y. Mu, J. Zhao, J. Feng, and B. Wang, "Characterizing Packet Loss in City-Scale LoRaWAN Deployment: Analysis and Implications," *IFIP Networking 2020 Conference and Workshops, Networking 2020*, pp. 704–712, 2020.

[124] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "On the Performance of Multi-Gateway LoRaWAN Deployments: An Experimental Study," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2020-May, 2020.

[125] M. C. Bor and T. Voigt, "LoRaSIM." [Online]. Available: https://mcbor.github.io/lorasim/ [Accessed: 20/05/2022]

[126] A. I. Pop, U. Raza, P. Kulkarni, and M. Sooriyabandara, "Does Bidirectional Traffic Do More Harm Than Good in LoRaWAN Based LPWA Networks?" *2017 IEEE Global Communications Conference, GLOBECOM 2017 - Proceedings*, vol. 2018-January, pp. 1–6, 2017.

[127] S. Li, U. Raza, and A. Khan, "How Agile is the Adaptive Data Rate Mechanism of LoRaWAN?" in *2018 IEEE Global Communications Conference, GLOBECOM 2018 - Proceedings*, 2018.

[128] G. Callebaut, G. Ottoy, and L. Van Der Perre, "Cross-Layer Framework and Optimization for Efficient Use of the Energy Budget of IoT Nodes," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2019-April, 2019.

[129] E. Sallum, N. Pereira, M. Alves, and M. Santos, "Performance optimization on LoRa networks through assigning radio parameters," *Proceedings of the IEEE International Conference on Industrial Technology*, vol. 2020-Febru, pp. 304–309, 2020.

[130] Sugianto, R. Harwahyu, A. A. Anhar, and R. F. Sari, "Simulation of mobile LoRa gateway for smart electricity meter," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 2018-Octob, pp. 292–297, 2018.

[131] F. Cuomo, J. C. C. Gamez, A. Maurizio, L. Scipione, M. Campo, A. Caponi, G. Bianchi, G. Rossini, and P. Pisani, "Towards traffic-oriented spreading factor allocations in LoRaWAN systems," *2018 17th Annual Mediterranean Ad Hoc Networking Workshop, Med-Hoc-Net 2018*, pp. 1–8, 2018.

[132] F. Cuomo, M. Campo, A. Caponi, G. Bianchi, G. Rossini, and P. Pisani, "Explora: Extending the performance of lora by suitable spreading factor allocations," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2017, pp. 1–8.

[133] F. Cuomo, M. Campo, E. Bassetti, L. Cartella, F. Sole, and G. Bianchi, "Adaptive mitigation of the Air-Time pressure in LoRa multi-gateway architectures," *arXiv*, pp. 56–61, 2019.

[134] Y. Oukessou, M. Baslam, and M. Oukessou, "LPWANIEEE 802.11ah and LoRaWAN capacity simulation analysis comparison using NS-3," *Proceedings of the 2018 International Conference on Optimization and Applications, ICOA 2018*, pp. 1–4, 2018.

[135] T. H. To and A. Duda, "Simulation of LoRa in NS-3: Improving LoRa Performance with CSMA," *IEEE International Conference on Communications*, vol. 2018-May, 2018.

[136] J. Finnegan, S. Brown, and R. Farrell, "Modeling the Energy Consumption of LoRaWAN in ns-3 Based on Real World Measurements," *2018 Global Information Infrastructure and Networking Symposium, GIIS 2018*, pp. 0–3, 2019.

[137] J. Finnegan, R. Farrell, and S. Brown, "Analysis and Enhancement of the LoRaWAN Adaptive Data Rate Scheme," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7171–7180, 2020.

[138] J. Finnegan, R. Farrell, and S. Brown, "Lightweight Timeslot Scheduling through Periodicity Detection for Increased Scalability of LoRaWAN," *Proceedings - 21st IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2020*, pp. 8–15, 2020.

[139] M. Slabicki, G. Premsankar, and M. Di Francesco, "Adaptive configuration of lora networks for dense IoT deployments," *IEEE/IFIP Network Operations and Management Symposium: Cognitive Management in a Cyber World, NOMS 2018*, pp. 1–9, 2018.

[140] A. M. Yousuf, E. M. Rochester, B. Ousat, and M. Ghaderi, "Throughput, Coverage and Scalability of LoRa LPWAN for Internet of Things," *2018 IEEE/ACM 26th International Symposium on Quality of Service, IWQoS 2018*, 2019.

[141] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," 2016.

[142] W. Weil, Ed., *Stochastic Geometry*, ser. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007, vol. 1892.

[143] O. Georgiou, C. Psomas, C. Skouroumounis, and I. Krikidis, "Optimal Non-Uniform Deployments of LoRa Networks," *IEEE Wireless Communications Letters*, vol. 2337, no. c, pp. 1–1, 2020.

[144] O. Georgiou, C. Psomas, and I. Krikidis, "Coverage Scalability Analysis of Multi-Cell LoRa Networks," *IEEE International Conference on Communications*, vol. 2020-June, 2020.

[145] N. Matni, J. Moraes, D. Rosario, E. Cerqueira, and A. Neto, "Optimal Gateway Placement Based on Fuzzy C-Means for Low Power Wide Area Networks," *Proceedings - 2019 IEEE Latin-American Conference on Communications, LATINCOM 2019*, 2019.

[146] H. Tian, M. A. Weitnauer, and G. Nyengele, "Optimized gateway placement for interference cancellation in transmit-only LPWA networks," *Sensors (Switzerland)*, vol. 18, no. 11, pp. 1–23, 2018.

[147] K. Q. Abdelfadeel, V. Cionca, and D. Pesch, "Fair Adaptive Data Rate Allocation and Power Control in LoRaWAN," *19th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2018*, pp. 6–14, 2018.

[148] J. Park, K. Park, H. Bae, and C. K. Kim, "EARN: Enhanced ADR with Coding Rate Adaptation in LoRaWAN," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11 873–11 883, 2020.

[149] S. Kim and Y. Yoo, "Contention-aware adaptive data rate for throughput optimization in LoRaWAN," *Sensors (Switzerland)*, vol. 18, no. 6, 2018.

[150] R. Piyare, A. L. Murphy, M. Magno, and L. Benini, "On-demand LoRa: Asynchronous TDMA for energy efficient and low latency communication in IoT," *Sensors (Switzerland)*, vol. 18, no. 11, pp. 1–22, 2018.

[151] B. Reynders, W. Meert, and S. Pollin, "Power and spreading factor control in low power wide area networks," *IEEE International Conference on Communications*, 2017.

[152] R. Marini, W. Cerroni, and C. Buratti, "A Novel Collision-Aware Adaptive Data Rate Algorithm for LoRaWAN Networks," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2670–2680, 2021.

[153] V. Hauser and T. Hegr, "Proposal of adaptive data rate algorithm for LoRaWAN-based infrastructure," *Proceedings - 2017 IEEE 5th International Conference on Future Internet of Things and Cloud, FiCloud 2017*, vol. 2017-Janua, pp. 85–90, 2017.

[154] T. Polonelli, D. Brunelli, and L. Benini, "Slotted ALOHA Overlay on LoRaWAN - A Distributed Synchronization Approach," *Proceedings - 16th International Conference on Embedded and Ubiquitous Computing, EUC 2018*, pp. 129–132, 2018.

[155] M. Rizzi, P. Ferrari, A. Flammini, E. Sisinni, and M. Gidlund, "Using LoRa for industrial wireless networks," *IEEE International Workshop on Factory Communication Systems - Proceedings, WFCS*, pp. 13–16, 2017.

[156] B. Reynders, Q. Wang, P. Tuset-Peiro, X. Vilajosana, and S. Pollin, "Improving reliability and scalability of LoRaWANs through lightweight scheduling," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1830–1842, 2018.

[157] R. Trüb and T. Lothar, "Increasing Throughput and Efficiency of LoRaWAN Class A," *Ubicomm*, no. c, pp. 54–64, 2018.

[158] A. Dongare, R. Narayanan, A. Gadre, A. Luong, A. Balanuta, S. Kumar, B. Iannucci, and A. Rowe, "Charm: Exploiting Geographical Diversity through Coherent Combining in Low-Power Wide-Area Networks," *Proceedings - 17th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN 2018*, pp. 60–71, 2018.

[159] R. Eletreby, D. Zhang, S. Kumar, and O. Yagan, "Empowering low-power wide area networks in urban settings," *SIGCOMM 2017 - Proceedings of the 2017 Conference of the ACM Special Interest Group on Data Communication*, pp. 309–321, 2017.

[160] G. F. Riley and T. R. Henderson, "The ns-3 Network Simulator," in *Modeling and Tools for Network Simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 15–34.

[161] D. Magrin, "Network level performances of a LoRa system," 2016.

[162] Signetlabdei, "LoRaWAN module." [Online]. Available: https://github.com/signetlabdei/lorawan [Accessed: 30/05/2022]

[163] 3GPP, "Radio Frequency (RF) system scenarios," *3GPP TR 36.942 V17.0.0*, 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2592 [Accessed: 30/09/2022]

[164] 3GPP, "Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)," *3GPP TR 45.820 V13.1.0*, 2015.

[165] S. S. Szyszkowicz, H. Yanikomeroglu, and J. S. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4222–4236, 2010.

[166] S. Schlegel, N. Korn, and G. Scheuermann, "On the interpolation of data with normally distributed uncertainty for visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2305–2314, 2012.

[167] Signetlabdei, "LoRaWAN module - Issue 98." [Online]. Available: https://github.com/signetlabdei/lorawan/issues/98 [Accessed: 01/06/2022]

[168] Signetlabdei, "LoRaWAN module - Issue 113." [Online]. Available: https://github.com/signetlabdei/lorawan/issues/113 [Accessed: 01/06/2022]

[169] Signetlabdei, "LoRaWAN module - Issue 105." [Online]. Available: https://github.com/signetlabdei/lorawan/issues/105 [Accessed: 1/06/2022]

[170] Signetlabdei, "LoRaWAN module - Issue 123." [Online]. Available: https://github.com/signetlabdei/lorawan/issues/123 [Accessed: 04/06/2022]

[171] B. Citoni, "My Modified LoRaWAN Module." [Online]. Available: https://github.com/brunocitoni/my-modified-lorawan-module [Accessed: 08/06/2022]

[172] The Things Network, "LoRaWAN airtime calculator."

[173] Avbentem, "Airtime calculator for LoRaWAN." [Online]. Available: https://avbentem.github.io/airtime-calculator/ttn/eu868 [Accessed: 01/06/2022]

[174] J. Huang, W. Susilo, and J. Seberry, "Observations on the message integrity code in ieee802. 11wireless lans," 2004.

[175] Signetlabdei, "SEM." [Online]. Available: https://github.com/signetlabdei/sem [Accessed: 08/06/2022]

[176] D. Magrin, D. Zhou, and M. Zorzi, "A Simulation Execution Manager for ns-3," in *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems - MSWIM '19*. New

York, New York, USA: ACM Press, 2019, pp. 121–125. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3345768.3355942 [Accessed: 04/05/2022]

[177] The Things Network, "The Things Network," https://www.thethingsnetwork.org/ [Online; accessed 12-May-2019].

[178] "IoT Scotland." [Online]. Available: https://iot-scotland.net/ [Accessed: 30/09/2022]

[179] NS-3, "Nakagami model." [Online]. Available: https://www.nsnam.org/doxygen/classns3_1_1_nakagami_propagation_loss_model.html [Accessed: 30/09/2022]

[180] R. Sanchez-Iborra, J. Sanchez-Gomez, J. Ballesta-Viñas, M. D. Cano, and A. F. Skarmeta, "Performance evaluation of lora considering scenario conditions," *Sensors (Switzerland)*, vol. 18, no. 3, 2018.

[181] F. Cuomo, M. Campo, E. Bassetti, L. Cartella, F. Sole, and G. Bianchi, "Adaptive mitigation of the Air-Time pressure in LoRa multi-gateway architectures," *arXiv*, pp. 56–61, 2019.

[182] A. Hoeller Jr, R. D. Souza, O. L. A. López, H. Alves, M. de Noronha Neto, and G. Brante, "Exploiting Time Diversity of LoRa Networks Through Optimum Message Replication," *Iswcs 2018*, pp. 1–5, 2018.

[183] B. Paul, "A Novel Mathematical Model to Evaluate the Impact of Packet Retransmissions in LoRaWAN," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 2020–2023, 2020.

[184] M. Sharif and A. Sadeghi-Niaraki, "Ubiquitous sensor network simulation and emulation environments: A survey," *Journal of Network and Computer Applications*, vol. 93, no. May, pp. 150–181, 2017.

[185] G. Fortino, C. Savaglio, G. Spezzano, and M. Zhou, "Internet of Things as System of Systems: A Review of Methodologies, Frameworks, Platforms, and Tools," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 223–236, 2021.

[186] The Things Network, "The Things Uno." [Online]. Available: https://www.thethingsnetwork.org/docs/devices/uno/ [Accessed: 30/09/2022]

[187] Semtech Corporation, "LoRaWAN – simple rate adaptation recommended algorithm," pp. 1–8, 2016.

[188] E. Harinda, S. Hosseinzadeh, H. Larijani, and R. M. Gibson, "Comparative Performance Analysis of Empirical Propagation Models for LoRaWAN 868MHz in an

Urban Scenario," *IEEE 5th World Forum on Internet of Things, WF-IoT 2019 - Conference Proceedings*, pp. 154–159, 2019.

[189] C. Caillouet, M. Heusse, and F. Rousseau, "Optimal SF allocation in LoRaWAN considering physical capture and imperfect orthogonality," *2019 IEEE Global Communications Conference, GLOBECOM 2019 - Proceedings*, 2019.

[190] C. Jiang, Y. Yang, X. Chen, J. Liao, W. Song, and X. Zhang, "A New-Dynamic Adaptive Data Rate Algorithm of LoRaWAN in Harsh Environment," *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–1, 2021.

[191] 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 *2015 14th International Conference on ITS Telecommunications (ITST)*, Dec 2015, pp. 55–59.

[192] P. Jörke, S. Böcker, F. Liedmann, and C. Wietfeld, "Urban channel models for smart city IoT-networks based on empirical measurements of LoRa-lmks at 433 and 868 MHz," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, vol. 2017-October, no. Cn I, pp. 1–6, 2018.

[193] H. Noura, T. Hatoum, O. Salman, J. P. Yaacoub, and A. Chehab, "LoRaWAN security survey: Issues, threats and possible mitigation techniques," *Internet of Things (Netherlands)*, vol. 12, 2020.