Exploitation ciblée du canal dans l'IoT : amélioration de la localisation et de l'allocation du spectre dans les réseaux LPWAN

Ahmed Abdelghany

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Télécommunications*

Par

Ahmed ABDELGHANY

Purposely Exploiting the Channel in IoT

Enhancement of Localization and Spectrum Allocation in LPWAN

Thèse présentée et soutenue à Rennes, le 08/12/2021
Unité de recherche : Univ Rennes, CNRS, IETR - UMR 6164
Thèse N° : 211552

Rapporteurs avant soutenance :

Laurent Clavier	Professeur, Institut Mines-Télécom (IMT)
Claire Goursaud	Maître de Conférences HDR, INSA Lyon

Composition du Jury :

Président :	Emmanuel Boutillon	Professeur, Université Bretagne-Sud (UBS)
Examinateur :	Benoît Denis	Chercheur, Chef de projets, CEA-Leti
Directeur de thèse :	Bernard Uguen	Professeur, Université de Rennes 1
Co-encadrant de thèse :	Dominique Lemur	Maître de Conférences, Université de Rennes 1

Invité(s) :

Christophe Moy	Professeur, Université de Rennes 1

Acknowledgement

This work would not have been possible without the help, guidance, and patience of my supervisor Bernard Uguen. I would like to warmly thank him for giving me this opportunity to conduct my Ph.D. under his supervision with the Propagation and Localization Group, in the Institute of Electronics and Telecommunications of Rennes (IETR), and for all the time he dedicated to help me. Many thanks to Dominique Lemur for his useful advice and constant support throughout the thesis period. Furthermore, I would like to thank Christophe Moy for the discussions we had during the thesis along with his encouragement, insightful comments what motivates me to succeed. I admit that my 3 years in the IETR lab was a great life experience mainly from their guidance and friendship attitude.

I wish to thank especially to my current and previous office mates Mor Diama, Jorge Ruiz and all fellows in the IETR lab for their help and kindness.

I am very grateful to my family for the unconditional support and enthusiasm that they gave me. My mother and father have always been my source of inspiration and they always motivated me to achieve better. I wish to thank my sisters Noha and Sana for their constant support, encouragement and understanding. And Finally, I would like to thank all my friends, especially Mohamed Atef and Adham Mahmoud, with whom I spent funny moments besides our daily interesting discussions. With their help and suggestions, they indirectly contributed to my well-being and this work.

Table of Contents

Abbreviations

2-D 2-Dimensional

ACK Acknowledgment

AD Amplitude Distance

ADR Adaptive Data Rate

AI Artificial Intelligence

AoA Angle of Arrival

AS Application Server

CARA Collision Avoidance Resource Allocation

CDF Cumulative Distribution Function

CFO Carrier Frequency Offset

CRC Cyclic Redundancy Check

CRLB Cramér-Rao lower bound

CSI Channel State Information

CSS Chirp Spread Spectrum

DQoC-A Discounted Quality of Channel Allocation

ESP Effective Signal Power

FEC Forward Error Correction

FIM Fisher Information Matrix

FIR Finite Impulse Response

GDoP Geometric Dilution of Precision

GNSS Global Navigation Satellite Systems

GPS Global Positioning System

IETR Institute of Electronics and Telecommunications of Rennes

IoT Internet of Things

ISM Industrial, Scientific and Medical

KNN K-Nearest Neighbors

LS Least Squares

LPWAN Low Power Wide Area Networks

LoRa Long Range

LoRaWAN Long Range Wide Area Network

LoS Line-of-Sight

LoB Lines-of-Bearing

LNS LoRa Network Server

MAB Multi-Armed Bandit

MAC Media Access Control

MAD Median Absolute Deviation

ML Maximum Likelihood

MUSIC MUltiple SIgnal Classification

NICs Network Interface Controllers

NLoS Non-line-of-sight

NSD Normalized Slope Distance

OSA Opportunistic Spectrum Access

OTAA Over-The-Air-Activation

PCC Pearson Correlation Coefficient

PDF Probability Density Function

PDR Packet Delivery Rate

PSD Power Spectral Density

QoC-A Quality of Channel Allocation

RBF Radial Basis Function

RF Radio Frequency

RFID Radio Frequency Identification

RMSE Root Mean Square Error

RSSI Received Signal Strength Indicator

Rx Receiver

SINR Signal-to-Interference-plus-Noise Ratio

SNR Signal-to-Noise Ratio

SF Spreading Factor

SFs Spreading Factors

SVM Support Vector Machine

TDoA Time Difference of Arrival

ToA Time of Arrival

Tx Transmitter

UCB Upper Confidence Bounds

UE User Element

USRP Universal Software Radio Peripheral

RÉSUMÉ EN FRANÇAIS

L'IoT a pris de l'ampleur au cours des dernières années en tant que nouveau paradigme permettant de connecter efficacement des milliards de dispositifs [1]. Au-delà des services traditionnels de voix, de vidéo et de données où le débit de données est l'objectif principal, dans le contexte du IoT, l'accent est mis dans ce mémoire sur les déploiements à faible coût avec de grandes zones de couverture. Pour fournir cette connectivité, les Low Power Wide Area Networks (LPWAN) sont considérés comme la principale technologie, en particulier le LoRaWAN dont la conception établit un compromis entre une faible consommation d'énergie et une grande portée de communication des dispositifs IoT [2], comme le montre la figure 1. Cette capacité se manifeste surtout dans les dispositifs LoRa typiques, qui peuvent couvrir des distances de plus de 10 km en utilisant une puissance d'émission maximale de 10 mW dans le spectre ISM sans licence, tout en maintenant des durées de vie de batterie extrêmement longues, jusqu'à plusieurs mois [3]. Ainsi, un module LoRaWAN fonctionnant dans la bande 868 MHz consomme 2.8 mA, à l'état "on", 38.9 mA en transmission de données et 14.2 mA en réception de données [4].

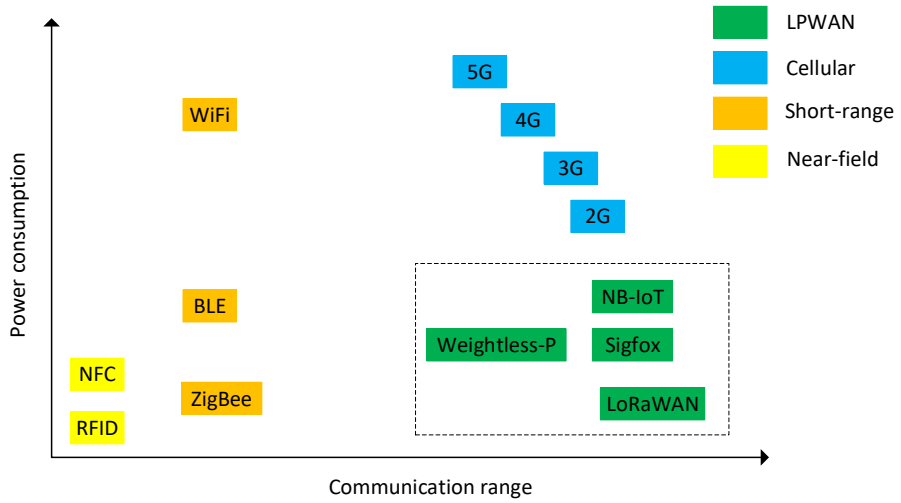


Figure 1 – Consommation d'énergie en fonction de la portée de communication pour divers protocoles radio, comme illustré dans [5].

La localisation est critique pour de nombreuses applications LPWAN en raison de la

nature même des données collectées à partir des terminaux [6]. Ces utilisations concernent les données de santé, la surveillance des animaux domestiques ou du bétail, de la faune ou des applications pour les villes intelligentes utilisant des capteurs tels que la température ou la surveillance de la qualité de l'air dans les environnements urbains [7]. Ce problème pourrait être résolu simplement en équipant chaque capteur d'une puce Global Navigation Satellite Systems (GNSS), par exemple en utilisant le Global Positioning System (GPS). Bien que cette solution soit facile, l'ajout d'un tracker GPS à un appareil augmente à la fois le coût et la consommation d'énergie [8]. Ainsi, il doit être rechargé tous les quelques jours car la consommation de courant d'un récepteur GPS est d'environ 30 mA à 50 mA, ce qui comprend à plus d'énergie que cette requise par la plupart des IoT à faible puissance. Par la suite, de nombreuses recherches dans le domaine de la localisation sans GPS dans les IoT, ont étudié des méthodes basées sur les RSSI, AoA, ToA, TDoA et leurs multiples intégrations [9]. Ces mesures sont utilisées par les passerelles pour déterminer leurs relations de position relative avec la source pour la localisation.

D'autre part, la perte de paquets sera le point faible de ces réseaux car elle dégrade les performances de l'ensemble du réseau à long terme. Par exemple, le protocole LoRaWAN ne disposant pas de droits exclusifs dans les bandes sans licence, également appelées bandes d'application ISM, des dégradations de canaux peuvent se produire sur la liaison entre le nœud final et la passerelle, réduisant ainsi la fiabilité des communications dans ces réseaux. Ces dégradations de canaux peuvent être dues à des collisions, comme le montre le document [10], ou à des effets de propagation, comme le montrent les documents [11], [12] et [13]. Cet échec de transmission peut avoir des conséquences graves pour diverses applications IoT, en particulier celles qui utilisent le mode de messagerie avec accusé de réception pour les données importantes des capteurs. Si le nœud final ne reçoit pas le paquet Acknowledgment (ACK), il retransmet le paquet de données. Cependant, cette retransmission nécessite une consommation d'énergie supplémentaire qui a un impact sur la durée de vie de la batterie des dispositifs du nœud final, en plus d'occuper une patron de spectre supplémentaire et d'augmenter le niveau d'interférence. De plus, un temps supplémentaire est occupé qui pourrait être utilisé pour les liaisons montantes. En outre, l'augmentation de la demande de retransmissions, en particulier du côté du réseau, entraînera l'épuisement des ressources énergétiques et une augmentation potentielle des émissions de CO_2. Par conséquent, l'industrie des télécommunications occupe une quantité équivalente de 2% à 10% de la consommation électrique mondiale. Cette quantité de

consommation d'énergie provient principalement des stations de base à hauteur de 60% à 80%.

Tout au long de cette thèse, la précision de la localisation est améliorée tout en préservant la faible consommation d'énergie du LPWAN, tandis que des techniques d'allocation de spectre sont proposées pour réduire les pertes de paquets. Ainsi, les propriétés de propagation du canal sans fil sont utilisées pour permettre aux dispositifs IoT de fonctionner avec une consommation d'énergie aussi faible que possible, comme indiqué dans les sections suivantes.

0.1 Objectifs et contributions

L'objectif initial de cette thèse est d'améliorer les techniques de localisation dans le réseau LPWAN qui dépendent de techniques centralisées (du côté du réseau). Ensuite, le travail s'oriente vers l'utilisation des mesures de canaux pour proposer des techniques d'allocation de spectre afin de réduire les pertes de paquets. En effet, ces deux approches utilisent les paramètres de propagation du canal pour réduire la consommation d'énergie dans le IoT afin d'être plus économe en énergie dans une perspective écologique, comme indiqué dans les sous-sections suivantes.

0.1.1 Localisation

La localisation est optimisée en améliorant la précision de la technique TDoA qui est considérée comme l'une des techniques les plus populaires, car elle ne nécessite aucune synchronisation du côté du nœud [4]. Seule la synchronisation temporelle du côté de la passerelle utilise un module GPS pour horodater le paquet reçu. Après avoir acquis les différences entre les horodatages d'une transmission, celle-ci est considérée comme un problème de multilatération qui implique la résolution d'un ensemble de fonctions hyperboliques. Ici, l'amélioration de TDoA est réalisée comme suit :

- Proposer une technique paramétrique de TDoA qui dépend des fonctions hyperboliques pour localiser le noeud sur une hyperbole, plutôt que de le localiser dans une position libre dans l'espace pouvant subir l'influence des imperfections de l'horodatage, comme le montre la figure 2.
- Un algorithme de prétraitement est proposé pour éliminer les horodatages aber-

rants à l'aide d'une méthode de seuillage robuste.

- Une autre méthodologie de prétraitement des TDoA pour éliminer les valeurs TDoA aberrantes est présentée, après avoir instrumenté un CRLB apparié. L'approche proposée détecte les meilleures valeurs de TDoA, qui ont les plus faibles valeurs de CRLB appariées, à proximité de l'emplacement du nœud supposé, en se basant sur une méthode de seuillage robuste.

Un simulateur est mis en œuvre pour l'analyse des performances de ces méthodes proposées dans un environnement de simulation bien défini. Ainsi, une comparaison est effectuée en étudiant les précisions atteignables pour la localisation basée sur ces approches TDoA proposées et la méthode classique TDoA, sur un environnement de simulation bien défini. Dans les résultats donnés, la faisabilité de ces techniques proposées est confirmée par une amélioration drastique sur une large gamme du nombre de passerelles ainsi que des variances du bruit de mesure. Comme le montre la Figure 3, la méthode paramétrique TDoA surpasse la méthode classique TDoA, ce qui indique que la méthode proposée est plus robuste aux fortes dérives des valeurs d'horodatage.

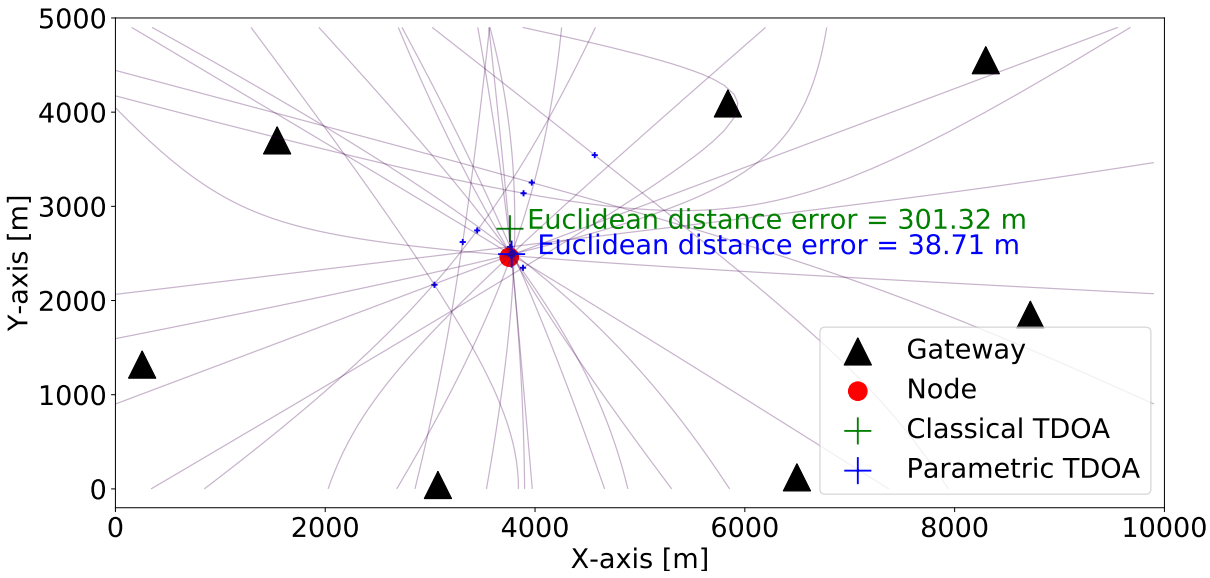
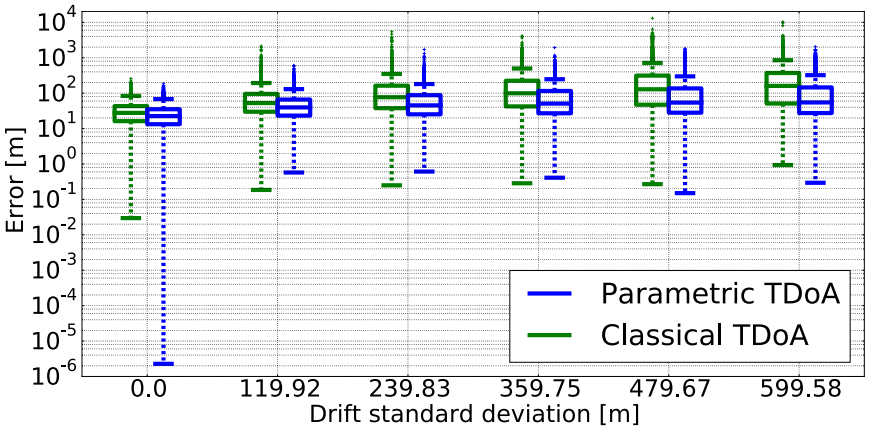
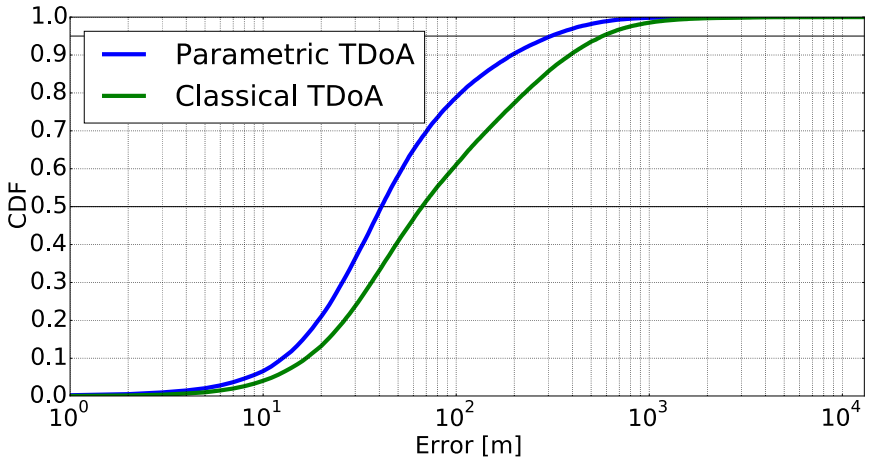


Figure 2 – Un scénario de localisation envisagé qui est généré avec le simulateur proposé.

D'autre part, la faisabilité de l'utilisation du CSI pour la localisation est étudiée, après avoir su que le CSI contient des informations sur le canal entre l'émetteur et le récepteur au niveau des sous-porteuses de données individuelles. Contrairement à la plupart des travaux précédents, l'approche envisagée consiste à utiliser l'intégralité des amplitudes des sous-porteuses sans calcul de moyenne ni réduction du CSI à bande étroite obtenu.



(a) Erreurs à chaque valeur de l'écart-type de la dérive.



(b) L'ensemble des résultats de simulation de σ_2 allant de 0 m to 600 m.

Figure 3 – Erreur de distance euclidienne et CDF tout en changeant l'écart type de la dérive σ_2.

De plus, le saut de fréquence dans les systèmes LoRa devrait être un avantage pour la localisation en permettant l'accès à une bande plus large. Afin d'obtenir une base fiable pour cette approche, une campagne de mesure en extérieur est réalisée dans la zone du Campus Beaulieu à Rennes pour estimer le CSI des signaux LoRa transmis depuis différents endroits. Les signaux LoRa transmis à partir de différents emplacements. Dans les résultats donnés, la faisabilité de l'utilisation de l'approche proposée est établie par la stabilité drastique de la pente du CSI dans le temps et l'espace, contrairement à l'amplitude moyenne du CSI. Ceci démontre la robustesse du CSI aux fluctuations du signal et sont rendu plus fiable que le RSSI.

0.1.2 Allocation du spectre

Le PDR est considéré comme une mesure majeure de la performance du réseau, tandis que sa dépendance aux dégradations du signal est confirmée [13]. Dans cette thèse, le PDR est étudié comme suit :

- Une étude approfondie de la dépendance en fréquence du PDR est effectuée en réalisant une campagne de mesure en extérieur dans la zone du Campus de Beaulieu à Rennes. A partir de chaque emplacement différent, les valeurs de ESP et Signal-to-Noise Ratio (SNR) sont obtenues ainsi que l'influence de ces paramètres sur le PDR est évaluée à chaque bande de fréquence indépendamment. Comme le montre la figure 4, la faisabilité de l'utilisation du ESP est prouvée par sa portée élargie lorsque le SINR est très faible, contrairement au RSSI qui présente une limitation.
- La modélisation du PDR à l'aide d'une fonction de distribution bêta ESP paramétrée est proposée. La faisabilité du modèle proposé est assurée par la simulation de PDR par rapport à l'ESP, comme le montre la figure 5. Ainsi, les valeurs de PDR simulées suivent bien la distribution des valeurs mesurées.

Cette investigation montre que l'utilisation de l'ESP dans les futures applications IoT, notamment la localisation à l'aide de l'empreinte digitale ESP. De plus, le modèle PDR proposé donne des directives importantes pour la régulation et l'optimisation des réseaux LoRaWAN futurs.

Sur cette base, la qualité du canal dans les différentes bandes de fréquences est prise en compte pour réduire les pertes de paquets dans le IoT, en particulier dans les réseaux LPWAN. Ainsi, les techniques d'allocation de spectre sont proposées comme suit :

- Une technique d'apprentissage QoC-A basée sur des algorithmes de bandit est pro-

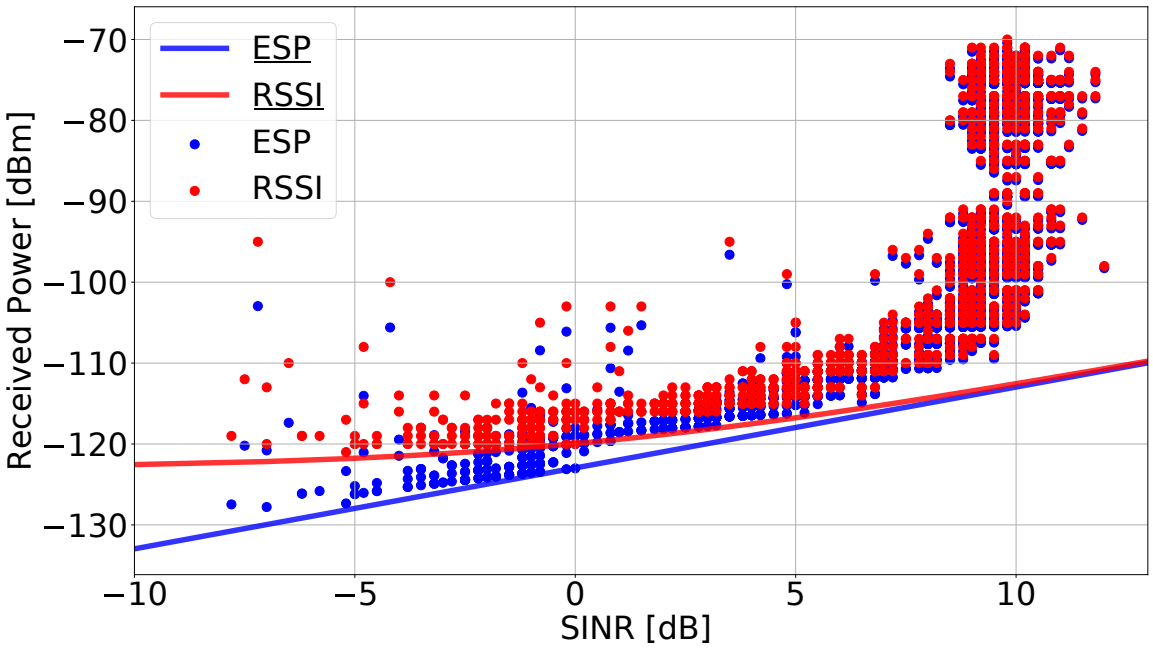
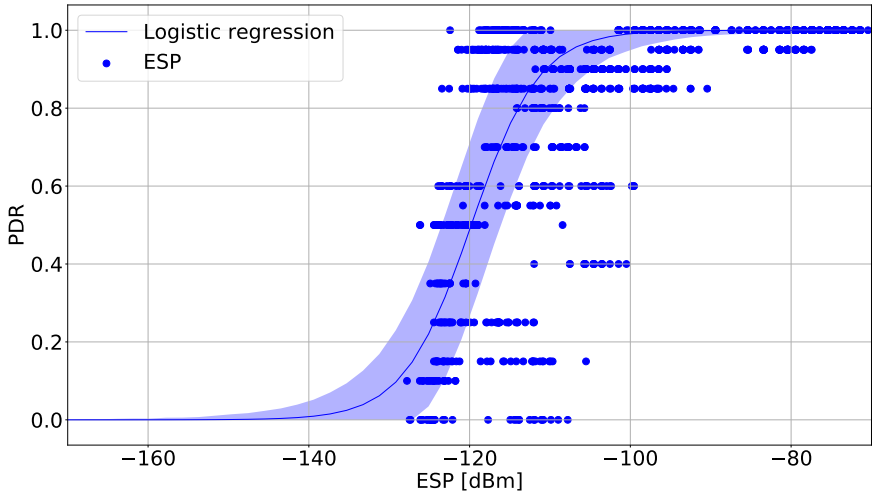


Figure 4 – Comparaison du RSSI et du ESP mesurés avec le SINR.

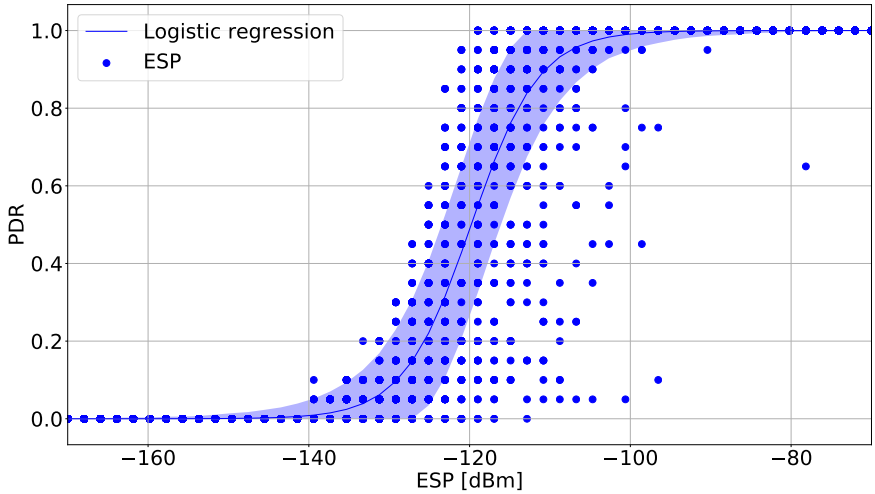
posée afin de choisir le canal de transmission, de sorte qu'il atténue les dégradations du canal en utilisant le paramètre de qualité du canal, c'est-à-dire ESP.

- Un algorithme DQoC-A est proposé pour s'adapter rapidement à tout changement abrupt dans les conditions des canaux.

De plus, une campagne expérimentale réaliste est réalisée dans la ville de Rennes pour démontrer la faible complexité et la faisabilité de ces algorithmes proposés tout en les implémentant du côté du dispositif IoT (approche décentralisée). Dans les résultats donnés, QoC-A surpasse la politique classique UCB avec un processus d'apprentissage plus accéléré. D'autre part, la faisabilité de l'utilisation de la DQoC-A dans les scénarios non stationnaires est prouvée par sa convergence rapide lorsque des changements abrupts dans les conditions des canaux se produisent. À la fin du processus, ces techniques d'apprentissage proposées sont assurées de donner moins de pertes de paquets que les techniques de pointe avec une allocation de fréquence aléatoire, comme le montrent le tableau 1 et le tableau 2. Ces algorithmes d'apprentissage par renforcement proposés sont recommandés pour être mis en œuvre du côté du nœud final. De plus, des remarques importantes sont données pour ajuster les configurations des algorithmes en fonction des propriétés de chacune des applications IoT potentielles.



(a) PDR mesuré en fonction de différentes valeurs de ESP.



(b) PDR simulé en fonction de différentes valeurs de ESP.

Figure 5 – PDR contre différentes valeurs de ESP.

Table 1 – Nombre de paquets transmis avec succès et perdus dans le scénario stationnaire

Algorithme	**Succès**	**Perte**
Uniform	668	132
UCB	761	39
QoC-A	768	32

Table 2 – Nombre de paquets transmis avec succès et perdus dans le scénario non-stationnaire

Algorithme	**Succès**	**Perte**
Uniform	405	195
UCB	487	113
QoC-A	497	103
DQoC-A	520	80

0.2 Organisation de la thèse

L'ordre de lecture du manuscrit peut être quelconque entre l'introduction du chapitre 1, et la Conclusion du dernier chapitre 7. La thèse est organisée en deux parties, correspondant aux deux colonnes de la figure 6.

Tout d'abord, dans la partie I (première colonne), elle commence par le chapitre 2 qui est nécessaire pour le reste de cette partie, car il présente l'état de l'art des techniques de localisation et les notations utilisées dans ce domaine. Les chapitres 3 présentent les approches proposées pour améliorer la précision des techniques TDoA. Ces améliorations se manifestent dans la méthode paramétrique TDoA proposée et les algorithmes de prétraitement. Dans le chapitre 4, la faisabilité de l'utilisation de la CSI pour la LPWAN localisation est affirmée expérimentalement.

Ensuite, la partie II (deuxième colonne) contient deux chapitres, qui peuvent être lus de manière indépendante. Le chapitre 5 commence par présenter les différents facteurs de pertes de paquets dans IoT, puis PDR est modélisé en fonction de l'ESP. Ainsi, il est prouvé que l'ESP est plus fiable que le RSSI sur la base d'une campagne de mesure réalisée. Enfin, le chapitre 6 présente les techniques d'allocation de spectre proposées pour réduire ces pertes de paquets et leurs résultats expérimentaux.

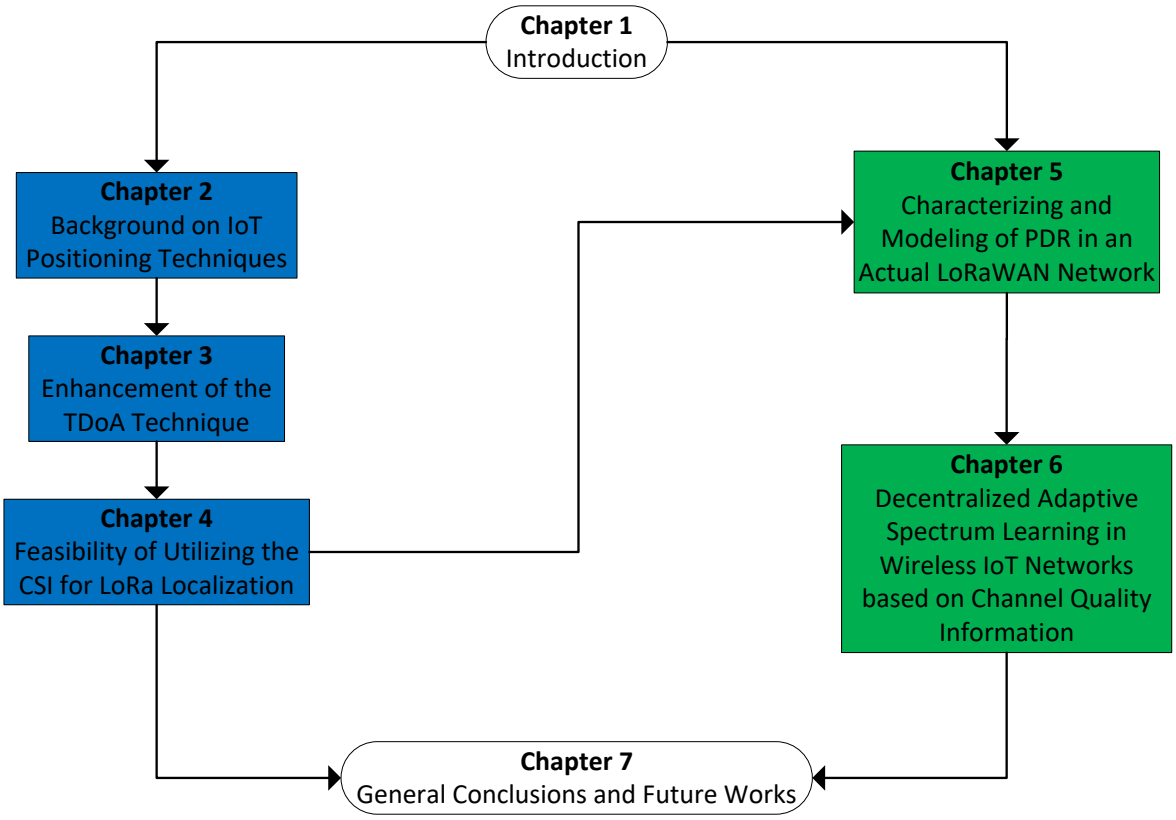


Figure 6 – Une carte de lecture de la thèse. Tout parcours descendant contenant le chapitre 1, au moins une des deux colonnes colorées et la conclusion est une manière possible de lire cette thèse.

Chapter 1

Introduction

IoT has been scaling up over the last few years as a new paradigm enabling the connection of billions of devices efficiently [1]. Beyond the traditional voice, video and data services where data throughput is the main purpose, in the context of IoT, the focus here is the low-cost deployments with large coverage areas. For providing this connectivity, LPWAN are considered the major technology, especially the LoRaWAN whose design compromises between low energy consumption and a large communication range of the underlying IoT-enabled devices [2], as shown in Figure 1.1. This ability is mostly exhibited in the typical LoRa devices, which can cover distances of more than 10 km using a maximum transmit power of 10 mW in the unlicensed ISM spectrum while maintaining extremely long battery lifetimes at the scale of up to several months [3]. Hence, a LoRaWAN module operating in the 868 MHz band consumes 2.8 mA, in the "on" state, 38.9 mA transmitting data and 14.2 mA receiving data [4].

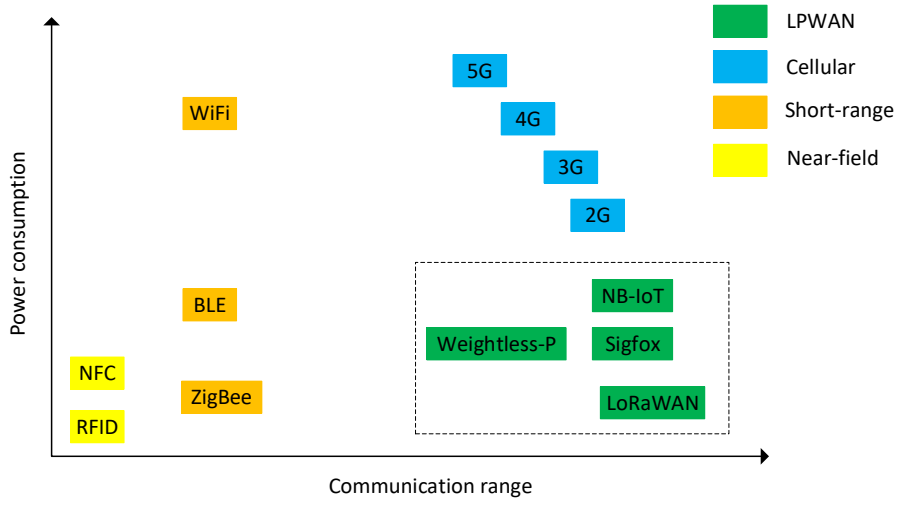


Figure 1.1 – Power consumption versus communication range for various radio protocols, as depicted in [5].

Localization is critical for many LPWAN applications due to the very nature of the

data collected from the terminals [6]. These usages involve health data, monitoring of pets or livestock, wildlife or applications for smart cities using sensors such as temperature or air quality monitoring in urban environments [7]. This problem could be simply solved by equipping each sensor with a GNSS chip, for example, using GPS. While this solution is tempting, adding a GPS tracker to a device will increase both cost and power consumption [8]. Thus, it needs to be recharged every few days as the current consumption of a GPS receiver is about 30 mA to 50 mA, which is considered more energy required by most low power IoT devices. Subsequently, plenty of researches within the domain of GPS-Free localization in IoT, have investigated methods based on RSSI, AoA, ToA, TDoA and their multiple integrations [9]. These measurements are utilized by the gateways to determine their relative position relations with the source for localization.

On the other hand, packet loss will be the weak point of these networks as it degrades the performance of the entire network in the long term. For instance, LoRaWAN protocol has no exclusive rights in the unlicensed bands [14], also called ISM application bands, channel impairments may occur on the link between the end node and the gateway, reducing the reliability of communications in these networks [15]. These channel impairments may occur either due to collisions as depicted in [10], or propagation effects as manifested in [11], [12] and [13]. Serious outcomes are caused by this transmission failure that may affect various IoT applications, in particular those using the acknowledged messaging mode for the important sensor data. If the end node does not receive the ACK packet, the end node will retransmit the data packet. However, extra energy consumption is required by this retransmission that impacts the battery life of the end node devices plus occupying an additional spectrum and raising the interfering level. Moreover, extra time is occupied that could be used for uplinks. Over and above, exhaustion of energy resources and a potential increase in the CO_2 emissions will occur by increasing the demand for these retransmissions, particularly at the network side. Hence, the telecommunication industry occupies an equivalent amount of 2% to 10% of the global power consumption [16]. This amount of power consumption is mainly from the base stations by an amount of 60% to 80% [17].

Throughout this thesis, localization accuracy is improved while preserving the low power consumption of the LPWAN, whereas spectrum allocation techniques are proposed to reduce the packet losses. Thus, the propagation properties of the wireless channel are

utilized to let the IoT devices function with less energy consumption as much as possible, as discussed in the following sections.

1.1 Objectives and Contributions

The initial target of this thesis is to enhance the localization techniques in the LPWAN that depend on centralized techniques (on the network side). Then, the work is shifted toward utilizing the channel measurements to propose spectrum allocation techniques owing to reduce the packet losses. Indeed, these two approaches are using the channel propagation parameters for reducing the power consumption in IoT to be more energy efficient in a green perspective, as briefed in the following subsections.

1.1.1 Localization

Localization is enhanced by improving the accuracy of the TDoA technique that is considered as one of the most popular techniques, as it does not need any synchronization on the node side [4]. Only time synchronization at the gateway side using a GPS module to timestamp the received packet. After acquiring the differences between the timestamps of a transmission, it is considered a multilateration problem that involves solving a set of hyperbolic functions. Here, the improvement in TDoA is achieved as follows:

- Proposing a parametric TDoA technique which depends on the hyperbolic functions to localize the node on a hyperbola, rather than locating it in a free position in the space potentially suffering from the influence of the timestamp imperfections.
- A pre-processing algorithm is proposed for dropping out the outlier timestamps based on a robust thresholding method.
- Another TDoA pre-processing methodology for dropping out the outlier TDoA values is presented, after instrumentalizing a paired CRLB. This proposed approach is detecting the best TDoA values, which have the lowest paired CRLB values specifically, in the vicinity of the guessed node location, based on a robust thresholding method.

A simulator is implemented for the performance analysis of these proposed methods in a well-defined simulation environment. Hence, a comparison is performed investigating the attainable accuracies for localizing based on these proposed TDoA approaches and the classical TDoA method, on a well-defined simulation environment. In the given results,

the feasibility of these proposed techniques is asserted by a drastic improvement over a wide range of the number of gateways as well as measurement noise variances. Moreover, the parametric TDoA outperforms the classical TDoA, hence, it indicates that the proposed method is more robust to the high drifts in the timestamp values.

On the other hand, the feasibility of utilizing the CSI for localization is studied, after knowing that the CSI contains information about the channel between the sender and receiver at the level of individual data subcarriers. Unlike most of the previous work, the intended approach is to use the entire subcarrier magnitudes without averaging or any reduction of the obtained narrowband CSI. Moreover, the frequency hopping in the LoRa systems should be a profit for localization by getting access to a wider band. In order to obtain a reliable basis for this approach, an outdoor measurement campaign is performed in the area of Beaulieu Campus in Rennes to estimate the CSI of transmitted LoRa signals from different locations. In the given results, the feasibility of using the proposed approach is asserted by the drastic stability of the CSI slope over time and space, contrary to the CSI average amplitude. This manifests the robustness of the CSI to the signal fluctuations and its more valuable rendering than the RSSI.

1.1.2 Spectrum Allocation

PDR is considered as a major measure of the network performance, whereas its dependency on the signal impairments is confirmed [13]. In this thesis, the PDR is investigated as follows:

- An in-depth investigation of the frequency dependency of the PDR is done by performing an outdoor measurement campaign in the area of Beaulieu Campus in Rennes. From each different location, the ESP and SNR values are obtained as well as the influence of these parameters on the PDR is evaluated at each frequency band independently. Within the given results, the feasibility of using the ESP is proven by its enlarged range when the SINR is very low, unlike the RSSI which has a limitation.
- Modelling the PDR using an ESP-parameterized beta distribution function is proposed. The feasibility of the proposed model is assured by simulating PDR against ESP, hence, the simulated PDR values follow the distribution of the measured ones well.

This investigation manifests using ESP in the future IoT applications, especially local-

ization using ESP fingerprinting. Moreover, the proposed PDR model gives important guidelines for future LoRaWAN network regulation and optimization.

Based on that, the channel quality across the different frequency bands is considered to reduce the packet losses in IoT, especially in LPWAN. Thus, spectrum allocation techniques are proposed as follows:

- A QoC-A learning technique based on bandit algorithms is proposed in order to choose the transmission channel, so that it mitigates the channel impairments using the channel quality parameter, i.e. ESP.
- A DQoC-A algorithm is proposed to adapt rapidly to any abrupt change in the channels' conditions.

Moreover, a real experimental campaign is performed in the city of Rennes to demonstrate the low complexity and the feasibility of these proposed algorithms while implementing them on the IoT device side (decentralized approach). In the given results, QoC-A outperforms the classical UCB policy with a more accelerated learning process. On the other hand, the feasibility of using the DQoC-A in the non-stationary scenarios is proven by its rapid convergence when abrupt changes in the channels' conditions occur. At the end of the process, these proposed learning techniques are assured to give fewer packet losses than the state-of-the-art ones with a random frequency allocation. These proposed reinforcement learning algorithms are recommended to be implemented in the end node side, moreover, important remarks are given for adjusting the algorithms' configurations based on the properties of each of the potential IoT application.

1.2 Organization of the Thesis

The reading order of the manuscript can be any top-down path between the Introduction in Chapter 1, and the Conclusion in the last Chapter 7. The thesis is organized in two parts, corresponding to the two columns of the following Figure 1.2.

First, in Part I (first column), it starts by Chapter 2 that is required for the rest of this part, as it introduces the state of the art localization techniques and the notations used in it. Chapters 3 present the proposed approaches for improving the accuracy of the TDoA techniques. These improvements are manifested in the proposed parametric TDoA method and the pre-processing algorithms. In Chapter 4, the feasibility of utilizing the

CSI for LPWAN localization is asserted experimentally.

Then, Part II (second column) contains two chapters, that should be read dependently. Chapter 5 starts by presenting different factors of the packet losses in IoT, subsequently, PDR is modeled as a function of ESP. Thus, ESP is proven to be more reliable than RSSI based on an accomplished measurement campaign. Finally, Chapter 6 introduces the proposed spectrum allocation techniques to reduce these packet losses and their experimental results.

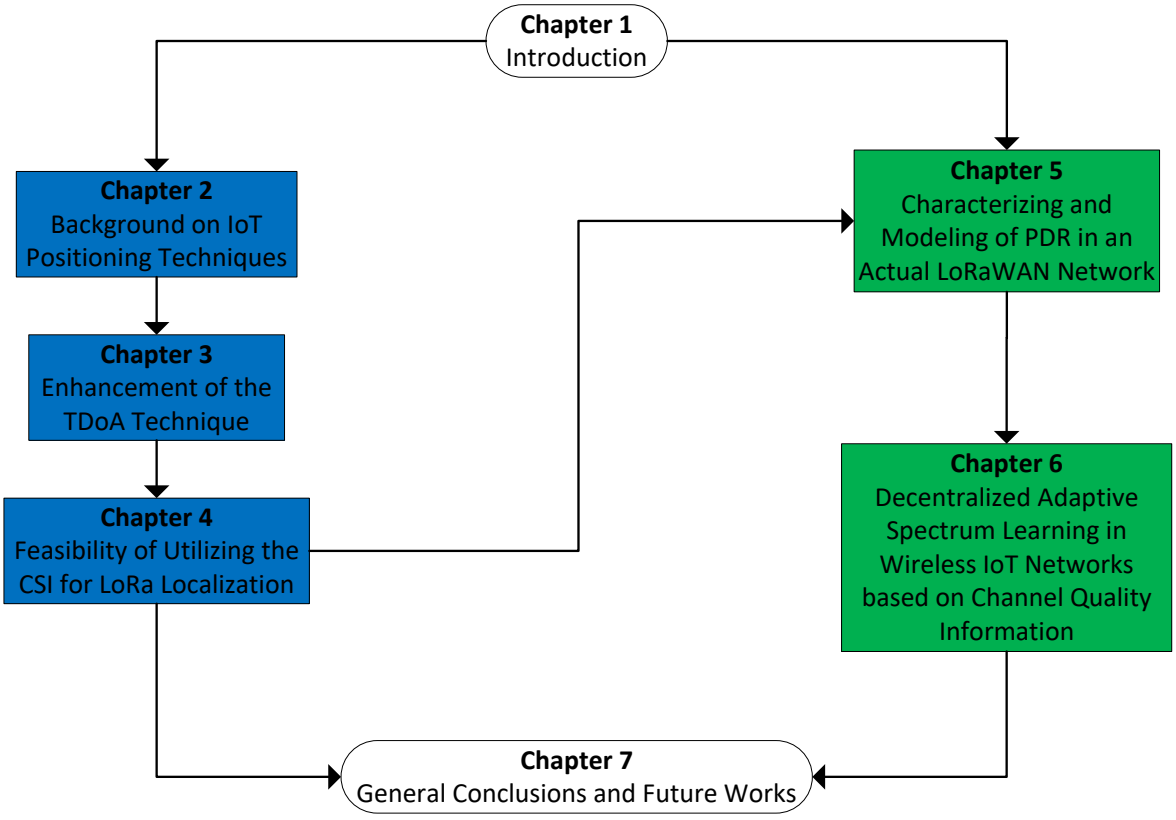


Figure 1.2 – A reading map of the thesis. Any top-down path containing Chapter 1, at least one of the two colored columns and the Conclusion is a self contained way to read this thesis.

1.3 Publications

This PhD thesis led to the publications that are mentioned in the following subsections.

1.3.1 Journals

1. **Ahmed Abdelghany**, Bernard Uguen, Christophe Moy and Dominique Lemur (2021), "Modelling of the packet delivery rate in an actual LoRaWAN network". Electronics Letters, 57: 460-462. https://doi.org/10.1049/ell2.12165
2. **Ahmed Abdelghany**, Bernard Uguen, Christophe Moy and Dominique Lemur, "Decentralized Adaptive Spectrum Learning in Wireless IoT Networks based on Channel Quality Information". The IEEE IoT Journal (IoT-J) 2021. (Submitted)

1.3.2 Conferences

1. **Ahmed Abdelghany**, Bernard Uguen and Dominique Lemur, "A Parametric TDoA Technique in the IoT Localization Context," 2019 16th Workshop on Positioning, Navigation and Communications (WPNC), 2019, pp. 1-6, doi: 10.1109/WPNC47567.2019.8970254.
2. **Ahmed Abdelghany**, Bernard Uguen and Dominique Lemur, "A Pre-processing Algorithm Utilizing a Paired CRLB for TDoA Based IoT Positioning," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020, pp. 1-5, doi: 10.1109/VTC2020-Spring48590.2020.9128385.
3. **Ahmed Abdelghany**, Bernard Uguen and Dominique Lemur, "A Robustness Comparison of Measured Narrowband CSI vs RSSI for IoT Localization," 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), 2020, pp. 1-5, doi: 10.1109/VTC2020-Fall49728.2020.9348854.
4. **Ahmed Abdelghany**, Bernard Uguen, Christophe Moy, Dominique Lemur, "On Superior Reliability of Effective Signal Power versus RSSI in LoRaWAN," 2021 28th International Conference on Telecommunications (ICT), 2021, pp. 1-5, doi: 10.1109/ICT52184.2021.9511510.

Part I

Radio Localization for IoT Infrastructures

Chapter 2

Background on IoT Positioning Techniques

Localization of the end devices is a mandatory requirement for specific usages, particularly in Industry 4.0 [6]. A GPS can achieve an accuracy of less than 10 m or even centimeter-level accuracy in an open outdoor scenarios [18]. However, GPS-based solutions are not feasible due to cost, processing and energy consumption, as discussed in the previous chapter. Hence, this chapter presents an overview of the localization techniques in the IoT, especially LPWAN which mainly depend on network-based localization methods [8]. Before proceeding, an overview of the LoRa technology is presented in Section 2.1. Then, Section 2.2 describes the main observable signal measurements that can be utilized for localization, as shown in Figure 2.1. Accordingly, Section 2.3 explains how to localize the end node using the two different types of algorithms that are based on a fingerprinting or geometric approaches.

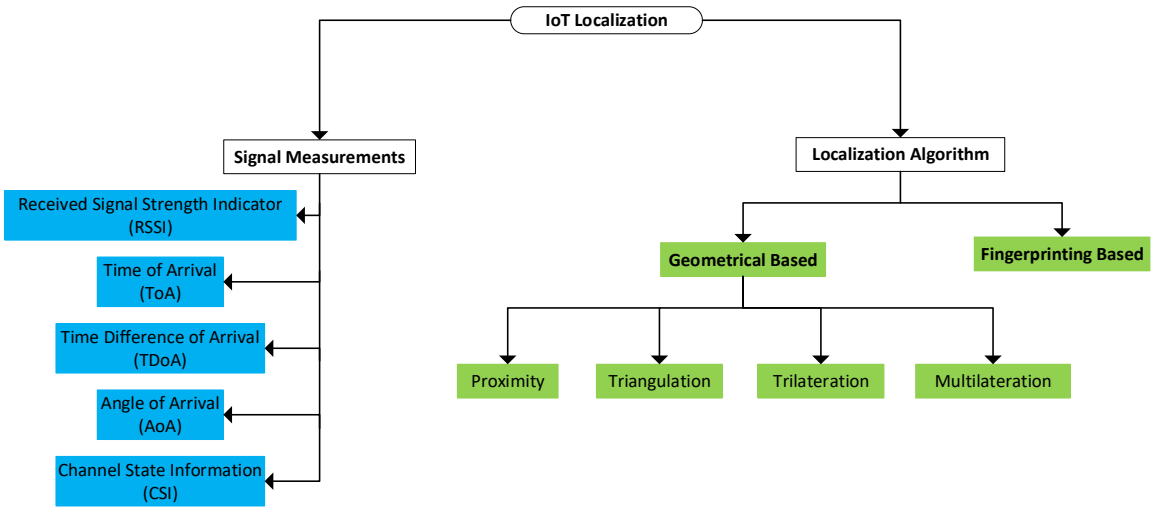


Figure 2.1 – Localization system in IoT including the signal measurements and classification of the positioning techniques.

2.1 LoRa Technology

This section first presents LoRa physical layer, then, the basic LoRaWAN network components are introduced.

2.1.1 LoRa Physical Layer

LoRa is considered the leading radio standard for wide area networks that is patented by the semiconductor manufacturer Semtech [19]. This technique uses a Chirp Spread Spectrum (CSS) modulation to improve robustness against noise and then allows Internet-connected applications to communicate with low-power devices over long-range wireless transmission [20]. Hence, LoRa symbols are encoded into multiple signals of increasing (up-chirp) or decreasing (down-chirp) frequencies. As shown in Figure 2.2, one LoRa symbol is generated from the m^{th} cyclic time shift of the basic chirp such as:

$$x_m[n] = \sqrt{\frac{1}{2^{SF}}} \exp(j2\pi \frac{(((m+n) \bmod 2^{SF}) - 2^{SF-1})^2}{2^{1-SF}}) \tag{2.1}$$

with $m = 0$ for the basic chirp symbol, while n depicts the sample index $n = 0, 1, 2, ..., 2^{SF} - 1$. Moreover, the Spreading Factor (SF) of the LoRa chirp can be modified ranging from 7 to 12 based on the required data rate and transmission range of the application [21]. Since one chirp consists of 2^{SF} samples, a higher SF increases the airtime of the packet and then resulting in greater receiving sensitivity. However, using high SF increases the power consumption and decreases the data rate. On the other hand, LoRa uses a bandwidth of 125 kHz, 250 kHz or 500 kHz to transmit a packet on the sub-1 GHz bands. In Europe, LoRa mostly operates in the 868 MHz ISM band.

Frame Structure

LoRa has a maximum packet size of 256 bytes [19], moreover, its frame structure contains the following main fields:

- **Preamble:** It is a variable-sized sequence of base chirps that is used for the time synchronization to detect the frame, as shown in Figure 2.2.
- **Frame synchronization symbols:** It can be used as a network identifier and it is composed of two modulated chirps [22].
- **Frequency synchronization symbols:** It is a two conjugate base chirps followed by a 0.25 conjugate chirp which are both utilized for fine frequency synchronization.

- **Header (optional):** It is a used field in the explicit mode that specifies the payload length, Forward Error Correction (FEC) code rate, the presence of a payload Cyclic Redundancy Check (CRC). In the implicit mode, this header field is not included to reduce the transmission time, hence, these configurations must be manually specified between the Transmitter (Tx) and the Receiver (Rx).
- **Payload:** It has the Media Access Control (MAC) payload containing the actual transmitted data. Moreover, its MAC header defines the frame type (data or ACK) and the protocol version. Finally, an optional 2-byte CRC of this data could be included.

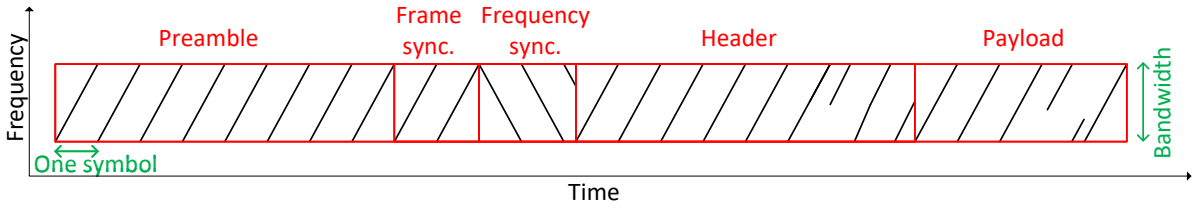


Figure 2.2 – LoRa packet structure.

2.1.2 LoRaWAN Network Architecture

LoRaWAN is a proposed MAC protocol by the LoRa Alliance for defining the network architecture and layers above the LoRa physical layer [2]. Without peering with a specific gateway, end nodes communicate with nearby gateways using an ALOHA-based protocol. Each gateway within range receives the message and then forwards it to the connected network server, as shown in Figure 2.3. Thus, the network server handles the message deduplication. On this basis, four basic elements compose a LoRaWAN network as follows:

1. **End Node:** It is an object with an embedded low-power communication device, hence, they are often battery operated. This can be a sensor, an actuator or both. Using LoRa modulation, these end devices are wirelessly connected to the LoRaWAN network through gateways. By default, the end node operates using class A, nevertheless, one of the three classes (A, B or C) can be used to imply a different media access procedure. Class A has two receiving windows scheduled to send the downlink packet in one of them, following each uplink packet. This class is the most commonly used for battery-powered devices, as it is considered the lowest energy consumption by spending most of the time in sleep mode. Class B devices

extend class A by adding scheduled receive windows for downlink messages from the server. Last, class C devices extend class A by continuously keep the receiving windows open unless they are transmitting a packet.

2. **Gateway:** It is an equipment with an antenna that receives broadcasts from the end nodes and sends data back to them. Each gateway is registered to the LoRaWAN network, subsequently, it forwards the messages to the network server through a backbone IP network based on a communication protocol. This communication protocol can be a cellular network, Wi-Fi, ethernet, fiber-optic or 2.4 GHz radio links.

3. **Network Server:** It is a server that manages the entire LoRaWAN network and routes messages from the end nodes to be delivered to the respective application server, and vice versa. It is also responsible for decoding the packets, managing Adaptive Data Rate (ADR), providing the Over-The-Air-Activation (OTAA) and acknowledgment of the confirmed data messages.

4. **Application Server (AS):** It is a server that hosts the final user application. This application allows users to interact or send data to the end nodes, moreover, measured data from sensors can be analyzed by it.

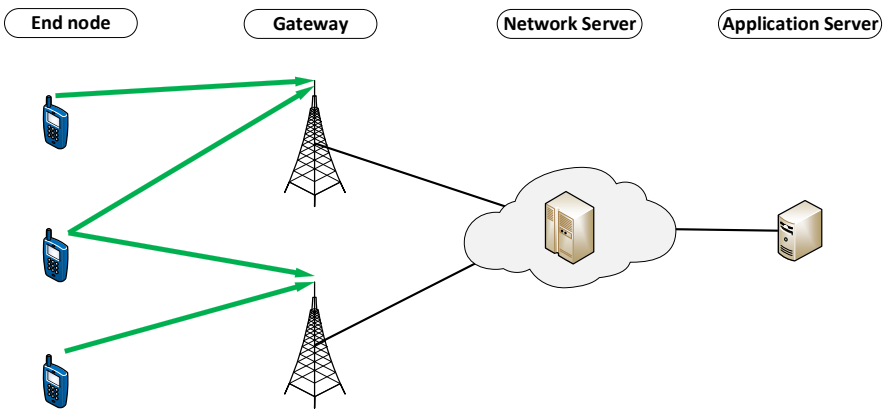


Figure 2.3 – General LoRaWAN architecture.

2.2 IoT Localization Signal Measurements

Passive geolocation is the preferred methods that depends on the information already available in most of the IoT networks. This section illustrates the commonly used localization signals that are estimated on the gateway side.

2.2.1 RSSI

Received Signal Strength Indicator (RSSI) is typically measured in dBm in the base station front end when the data packet is received from the end node [23, 24], as shown in Figure 2.4. In free space, this value directly reflects the distance between the Tx and Rx [25]. The high RSSI value obtained means that Tx and Rx are close to each other and vice versa. In practical life, the distance d between Tx and Rx can be determined using different path loss models where the transmission power at the Tx is known. For example, the determined RSSI value is related to a distance d using the following path loss model [26]:

$$RSSI = C - 10n \log_{10}(d), \tag{2.2}$$

where C is a fixed constant that accounts for system losses, n is the path loss exponent that varies depending on the environment. Thus, the value of n determines how quickly RSSI falls with respect to d.

Using RSSI has many advantages as:

- No extra hardware is required on the gateway side, hence, RSSI can be straightforwardly obtained [27].
- RSSI can be flexibly used for various localization algorithms, such as trilateration or fingerprinting approaches [28], as detailed in the following section.

However, the challenges for using RSSI include:

- The path-loss model parameters are difficult to be determined accurately in many scenarios such as urban and indoor environments [29, 30].
- RSSI value variations and interference often happen, especially over large transmission distances, due to environmental factors [31, 23]. These uncertainties are problematic issues as every 1 dB change in the RSSI value may lead to distance differences of meters.

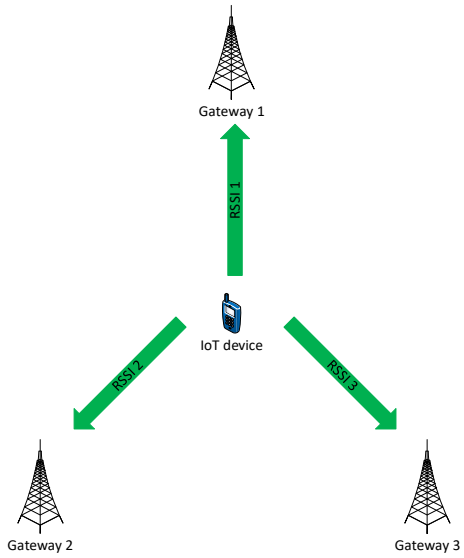


Figure 2.4 – Example of localization based on RSSI.

2.2.2 ToA

ToA is defined as the time spent by the signal to propagate between the end node and the gateway [32]. The distance d is acquired simply as the product of the measured transmission time by the speed of light ($c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$) as:

$$d = c \times (t_{arrived} - t_{transmitted}), \tag{2.3}$$

where $t_{arrived}$ and $t_{transmitted}$ are the time of receiving and transmitting the packet, respectively.

Using ToA includes the following advantages:

- The distance d can be obtained directly by linearly converting the ToA value to a spacial distance without knowing any path loss model, as shown in Figure 2.5.
- Localizing with ToA can achieve high accuracy of centimeter-level in Line-of-Sight (LoS) environments [33].
- The accuracy assessment of ToA method is available [34].

Nevertheless, ToA localization requires precise timing on both the end node and gateway, or synchronization between them. For instance, ten-nanosecond-level timing accuracy is required to achieve meter-level ranging. Such precise timing is not affordable for many IoT devices with low complex hardware.

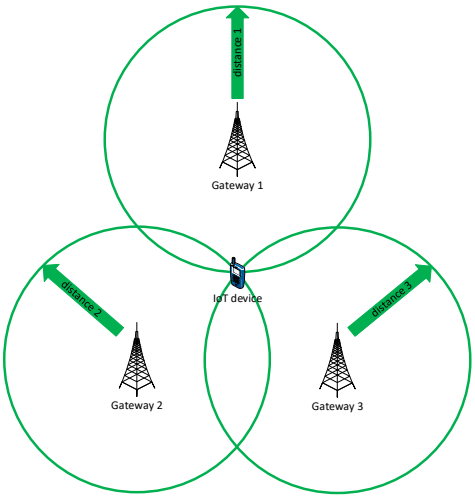


Figure 2.5 – Example of localization based on trilateration.

2.2.3 TDoA

TDoA is measured by computing the signal arrival time differences among multiple gateways [35]. As shown in Figure 2.6, a mobile node transmits data to the network. Each gateway within reach records the timestamp of the received packet. The timestamps of each gateway are then forwarded to the network server which in turn sends a request to the geolocation solver.

Hence, TDoA localization has the following advantages:

- No precise timing on nodes is required. Instead, only the gateways are required to have synchronized clocks. This accurate time synchronization can be achieved easily by the use of a GPS receiver at each gateway.
- An accurate localization of hundreds to thousands of meters can be achieved based on TDoA localization methods.
- TDoA localization methods have many accuracy assessment mechanisms in theory [36].

On the other hand, the challenges for TDoA localization include:

- TDoA localization accuracy is affected by the signal bandwidth W [37]. In case of multi-path, a receiver can recognize a minimum difference in the distance Δd that is traveled by two signals as:

$$\Delta d = \frac{c}{W}, \tag{2.4}$$

where c is the speed of light. A direct and a reflected signal with a path length difference less than Δd will not be distinguished as separate signals but grouped as these signals will be summed and the phase shift will be added. This will result in harming the range estimation. For the typical LoRa signal with bandwidth $W = 125\,\text{kHz}$, this gives the worst-case ambiguity as $\pm\frac{\Delta d}{2} \approx \pm 1200\,\text{m}$ [38].

- To have an accurate time synchronization, the base stations' cost increases by adding a GPS receiver module [39].
- The TDoA based localization algorithms are more sensitive to the noise [40]. Moreover, the impact of noise is enhanced by the use of differential measurements.

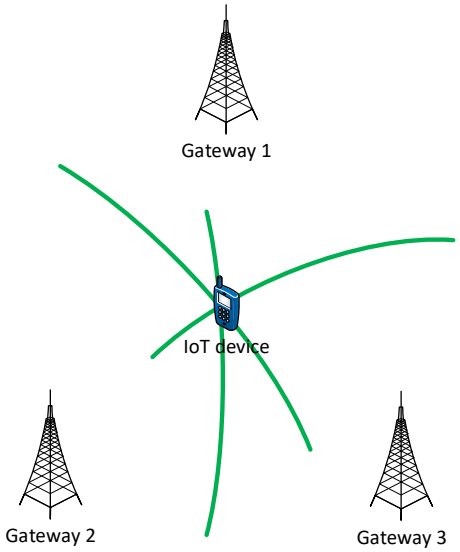


Figure 2.6 – Example of localization based on TDoA.

2.2.4 AoA

The AoA estimation is a mechanism that indicates the direction of arrival of the received signal by processing the incoming signal on an antenna array [41], as shown in Figure 2.7. Thus, AoA positioning has the following advantages:

- A high accuracy of a decimeter level can be achieved using a typical AoA localization system [42].
- Few gateways are needed to estimate a position [42]. For 2-Dimensional (2-D) localization, using two gateways only is feasible which is less than the required in ToA or TDoA.

- The well researched accuracy assessment mechanism of the AoA localization approaches [43].

However, the challenges for AoA localization include:

- Relatively large and complex hardware of the AoA system as it requires specific hardware, such as an antenna array and phase detector [44]. These requirements make the IoT gateway has a high cost.
- Both angular measuring and positioning accuracy degrade significantly by shadowing and multipath reflections [45, 46, 47], particularly when the distance between the node and gateway increases in LoRaWAN networks [48].

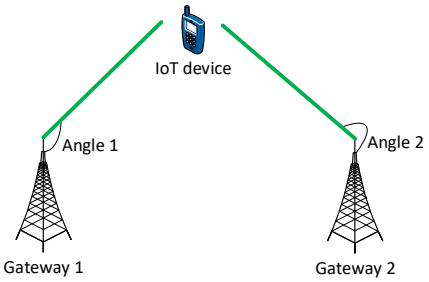


Figure 2.7 – Example of localization based on AoA.

2.2.5 CSI

CSI between IoT node and the gateway is possible to be obtained to identify each location by it [49]. Moreover, utilizing CSI for localization has the following advantages:

- A decimeter-level or higher accuracy can be achieved, especially for wide-band signals such as the typical Wi-Fi signal with bandwidth $W = 40\,\text{MHz}$ [50].
- CSI has more features than RSSI, thus, CSI is fine-grained information from the physical layer which describes the amplitude and phase on each sub-carrier in the frequency domain [51]. Hence, the signal travels along different fading or scattering paths on account of the multi-path effects over different sub-carriers. In the frequency diversity attribute of CSI, these bring different amplitudes and phases on each sub-carrier.
- CSI is more robust to noise and signal strength fluctuations which contribute to most of the estimation errors in the localization systems [52, 53]. Hence, RSSI is easily affected by the temporal and spatial variance due to the multi-path effect.

However, not all the Network Interface Controllers (NICs) or the IoT gateways provide the CSI information [54].

2.3 IoT Localization Methods

This section presents the main state-of-the-art localization approaches with the advantages and challenges for each type of localization method. Importantly, these approaches don't impose any hardware or firmware changes for the end devices. Actually, these algorithms convert the different signal properties between the end node and the gateway to estimate the position of the node [55]. Before the development of the machine learning techniques, the majority of the localization methods were based on geometric measurements, such as distances and angles. Later, Database-Matching (fingerprinting) methods are developed, thanks to the machine learning techniques and the diversification of the signal measurements, as discussed in the following subsections.

2.3.1 Geometrical Localization Methods

Range-based techniques are based on the measurement of distances or angles between the end node and the gateway, as the following detailed techniques.

Proximity

Proximity is regarded as a simple localization approach as it considers the estimated node location as the gateway location [56]. Therefore, it is commonly used in localization systems based on Radio Frequency Identification (RFID) and infrared [57]. Moreover, the cell identification system of the mobile phone is another example of proximity. This method detects the probable position of a User Element (UE) by identifying which cell id is used at a given time from which base station, whereas the base station location is well-known by the cellular network. Consequently, the proximity technique does not provide an absolute position of an object, only a piece of location information is provided [58]. Hence, the gateway with the highest RSSI is considered to estimate the node position when the transmitted packet is identified by more than one gateway [59]. On this basis, the performance of this approach relies on the dense deployment of the gateways with well-known positions.

Although this coarse localization method has the largest localization uncertainty, it has the lowest computational load and the lowest number of gateways. Subsequently, this method is profitable for bridging the outages when the other localization methods do not have a sufficient number of gateways or signal measurements are acquired with low precision.

Triangulation

The triangulation technique is the way to use the geometric properties of triangles to localize the targeted nodes [44]. It estimates the node location by computing the angles relative to multiple gateways, as shown in Figure 2.7. Using at least two gateways' locations as references, the intersection of the two independent Lines of Bearing (LoB) gives the estimated 2-D-position. A set of N gateways are considered with known positions

$$\mathbf{G} = \Big[\mathbf{g}_1, \ldots, \mathbf{g}_n, \ldots, \mathbf{g}_N\Big] \tag{2.5}$$

with

$$\mathbf{g}_n = [x_n, y_n]^T, \tag{2.6}$$

while $\mathbf{p} = [x, y]^T$ is any node position in the plane. Thus, the measured angle of the received signal from $\mathbf{p}$ on gateway n is defined as:

$$\hat{\theta}_n = \tan^{-1}\left(\frac{y - y_n}{x - x_n}\right) + u_n, \tag{2.7}$$

where $u_n \sim \mathcal{N}(\delta, \sigma^2)$ is an additive Gaussian random variable which is assumed independent and identically distributed (*i.i.d.*). In the classical AoA localization techniques, the node location $\hat{\mathbf{p}} = [\hat{x}, \hat{y}]^T$ is estimated using the conventional algorithms such as the Least Squares (LS), or by minimizing a cost function on the unknown coordinates as:

$$\hat{\mathbf{p}} = \arg\min_{x,y}\Big(\sum_{i=1}^{N} |\hat{\theta}_i - \tan^{-1}\left(\frac{y - y_i}{x - x_i}\right)|\Big). \tag{2.8}$$

For example, a node positioning method is experimentally proposed based on AoA estimation in [42]. For the angle estimation, MUltiple SIgnal Classification (MUSIC) algorithm is used due to its high angular resolution and sensitivity. The AoA is estimated

at two gateways (anchors) which are equipped with 4 antennas. For the defined scenario in Figure 2.8, this approach is able to successfully determine the position of a node with an average accuracy of 14 cm.

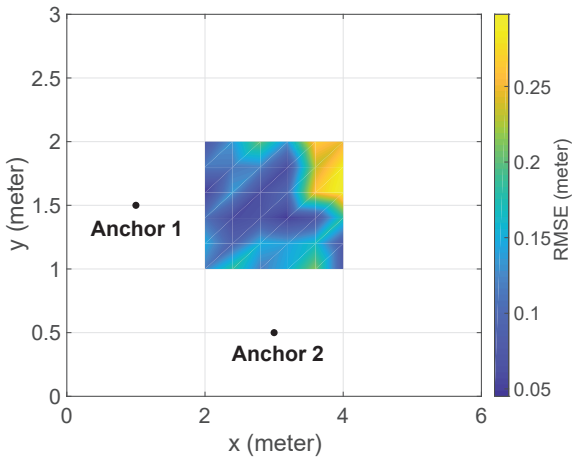


Figure 2.8 – RMSE of the estimated node position [42].

Trilateration

The trilateration technique estimates the node position using the geometric properties of circles and spheres [60], as shown in Figure 2.5. This classical technique is widely used in GNSS as well as indoor positioning. In contrary to the triangulation, at least three gateways at known reference locations are used to estimate the 2-D position of the targeted node. Besides, the node altitude can be obtained as well as precision of the node location can be improved using more distance measurements. These distance measurements can be obtained from the ToA values or by converting the RSSI values into distances using a path loss model as in [23, 61, 62]. Using each distance measurement $\hat{d}$ from a gateway, a circle of ambiguity, where the node can be located, is defined. Accordingly, the node location can be estimated from the intersection of these circles using the Pythagorean theorem as:

$$\hat{\mathbf{p}} = \arg\min_{x,y}\left(\sum_{i=1}^{N} |\hat{d}_i - \sqrt{(x-x_i)^2 + (y-y_i)^2}|\right). \tag{2.9}$$

On this basis, [63] develops a GPS-independent positioning system using a low-cost LoRa device. Using the trilateration algorithm, an RSSI-based positioning system is devel-

oped and evaluated on different rural scenarios and environmental considerations. Using this low-cost system, an average error of ≈12.11 m is obtained for a 150 m × 150 m scenario.

Multilateration

Multilateration is solving a set of hyperbolic functions to localize the node after acquiring the differences between the timestamps (TDoA values) of a transmission [64]. Therefore, at least three gateways are needed to locate the node on the intersection point of the hyperbolas in 2-D space, as shown in Figure 2.6. Indeed, by increasing the number of gateways, more hyperbolas are acquired, and then the higher the accuracy is obtained with denser intersections at the node location [65]. First, the timestamp of the transmitted packet from the node location $\mathbf{p}$ is measured on gateway n as:

$$\tau_n = \frac{d_n}{c} + u_n, \tag{2.10}$$

where c is the celerity of light and $u_n \sim \mathcal{N}(\delta, \sigma^2)$ is an additive Gaussian random variable which is accounting for the departure between the timestamp τ_n and the node to gateway time of flight toa_n. Notice that the standard deviation σ could be made dependent on the gateway index but this dependency is here omitted. Moreover, the time offset δ is assumed to be equal for all the gateways, and LoS conditions are implicitly assumed, while $d_n = \|\mathbf{g}_n - \mathbf{p}\|_2$ is the distance between any node position and the gateway n, where $\|.\|_2$ denotes the 2-norm. Each combination of 2 among N gateways leads to:

$$H = \binom{N}{2} = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2} \tag{2.11}$$

constraints, each being associated with one hyperbola. An hyperbola j involves the 2 gateways $l(j)$ and $r(j)$, while $\mathbf{l} = \Big[l(1), \quad l(2), \quad \dots \quad , l(H)\Big]$ and $\mathbf{r} = \Big[r(1), \quad r(2), \quad \dots \quad , r(H)\Big]$. Accordingly,

$$\mathbf{G_r} = \Big[\mathbf{g}_{r(1)}, \dots, \mathbf{g}_{r(H)}\Big], \tag{2.12}$$

$$\mathbf{G_l} = \Big[\mathbf{g}_{l(1)}, \dots, \mathbf{g}_{l(H)}\Big], \tag{2.13}$$

$$h_{l(i),r(i)} = d_{l(i)} - d_{r(i)}, \tag{2.14}$$

$$\tau_{l(i),r(i)} = \tau_{l(i)} - \tau_{r(i)}, \tag{2.15}$$

$$\mathbf{h} = \left[h_{l(1),r(1)}, \ldots, h_{l(H),r(H)}\right] \tag{2.16}$$

and

$$\boldsymbol{\delta}_\tau = \left[\tau_{l(1),r(1)}, \ldots \tau_{l(H),r(H)}\right]. \tag{2.17}$$

In the classical TDoA techniques, the estimated node location $\hat{\mathbf{p}} = [\hat{x}, \hat{y}]^T$ is obtained using the conventional algorithms such as the LS or under Gaussianity assumption by Maximum Likelihood (ML) estimation [66]:

$$\hat{\mathbf{p}}_{ML} = \arg\min_{x,y}((\mathbf{h} - c\boldsymbol{\delta}_\tau)^T \mathbf{R}^{-1}(\mathbf{h} - c\boldsymbol{\delta}_\tau)) \tag{2.18}$$

with the associated $H \times H$ covariance matrix:

$$\mathbf{R} = \begin{bmatrix} 2\sigma^2 & \sigma^2|0 & \ldots & \ldots \\ \sigma^2|0 & 2\sigma^2 & \ldots & \ldots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ldots & 2\sigma^2 \end{bmatrix}, \tag{2.19}$$

where the notation $\sigma^2|0$ means that each off-diagonal term of $\mathbf{R}$ is either equal to σ^2 or 0 provided that the two couples of gateways i.e., involved in that term, share respectively either one gateway or none.

Originally, multilateration is utilized for military purposes such as aircraft monitoring and detecting jammer signal sources in the 1960s [67]. Later, it is used in the control towers in the airports to detect the airplane position passively without the need to send it a special request. In case GPS does not function as intended, this is considered a backup localization approach. With the rising star of IoT, multilateration is often preferred as one of the localization methods by many LPWAN network providers, after considering the cost of implementing localizations and the limits of these networks [68].

In [4], the authors implement a localization system that is capable of estimating the

node location by exploiting the transmitted LoRa packets. This is done by applying the multilateration algorithm on the gateway timestamps of the received packet. This experiment is accomplished outdoor in Denmark using only four gateways, a server and a java application to store the obtained information in a database. The results demonstrated that it is feasible to localize a stationary node with an accuracy of ≈100 m using a TDoA based technique (multilateration).

Furthermore in [69], the performance of TDoA geolocation for outdoor tracking purposes has been investigated on a public LoRaWAN network. For different trajectories (walking, cycling, and driving) and LoRa Spreading Factors (SFs), localization accuracy, update probability, and update frequency are evaluated. The results show a median localization accuracy of 200 m, whereas 90% of the obtained errors are less than 480 m. When varying the mobility or the SF value, no trend is found on the performance of the localization error but in general, the best localization accuracies are obtained for SF12. Hence, a higher probability of a location update is a result of using a higher SF with a higher communication range. On the other hand, the number of potential location updates can be increased when using lower SF with more transmissions allowance after respecting the duty cycle.

2.3.2 Fingerprinting Localization Methods

Fingerprinting techniques utilize the database of previously measured signal parameters corresponding to well-known node positions [70]. Thus, the fingerprinting approaches find the closest matched location by computing the difference between the measured fingerprint and the reference fingerprint in the database [71]. In contrary to the ranging methods, the gateways' locations are not required in order to estimate the location of the end node. Consequently, this localization process consists of two phases as follows:

1. **Offline (training) phase:** Signal features are extracted from known positions [72]. These signal measurements are considered as the training data which can be RSSI, CSI, magnetic intensity, visual features or any other signal properties [73]. In this thesis, RSSI values are estimated from different locations using LoRaWAN field test device, as shown in Figure 2.9. Accordingly, the Radial Basis Function (RBF) interpolation method is utilized for interpolating the RSSI values and then producing the heat map.

2. **Online phase:** The real-time measured signal feature is compared with the stored database to predict the node location [74]. To map the measured feature to the reference one, many pattern recognition techniques are commonly used such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and neural networks [75].

To have an accurate localization, the signal measurements, i.e. acquired in the training phase, should be stable over time for each location. However, the environment can alter rapidly, moreover, the signal measurements can be affected by thermal noise, multi-path, scattering and reflection effects [76, 77]. On the other hand, significant time and hard work are required in the offline stage to generate and update the database or filter out the outlier signal measurements [78, 79, 80]. Nevertheless, fingerprinting-based localization techniques are often better than the geometrical ones in terms of performance and localization accuracy [81].

In [82], the usability and accuracy of the fingerprinting localization approach are evaluated for the large coverage areas in the LoRaWAN networks. In a 340 m × 340 m test area, a private LoRa network is constructed using four gateways and an end node to collect the training RSSI measurements in the offline phase. Consequently, the RBF interpolation method is utilized for interpolating the RSSI values to fill grids that have no RSSI value in the matrix with 340 × 340 cells. After fully interpolating this RSSI map, the location is estimated using proposed probabilistic methods in the online phase. Based on that, a mean localization error of 24.1 m is obtained using 46 test points.

On the other hand, the work in [83] is focused on averting the time and effort that are consumed for acquiring manually the signal measurements in the offline phase. For that, SateLoC technique is proposed that uses the high-resolution satellite images of the area of interest to generate the fingerprinting map. Thus, the environment type (such as urban, rural or vegetation) of each pixel in the image is identified and then assigned to a proper path loss model, as shown in Figure 2.10. To obtain the expected signal power at each pixel, an overall path loss is calculated that accounts for all the traversed environment types of an arbitrary LoRa link (from the gateway to the targeted pixel), as shown in Figure 2.11. This virtual fingerprinting approach is evaluated by localizing commercial LoRa devices in an urban area of 277 500 m^2. The results confirm that this approach outperforms the

classical model-based approaches by achieving a median localization error of 47.1 m, while saving the human effort during the fingerprinting acquisition process.

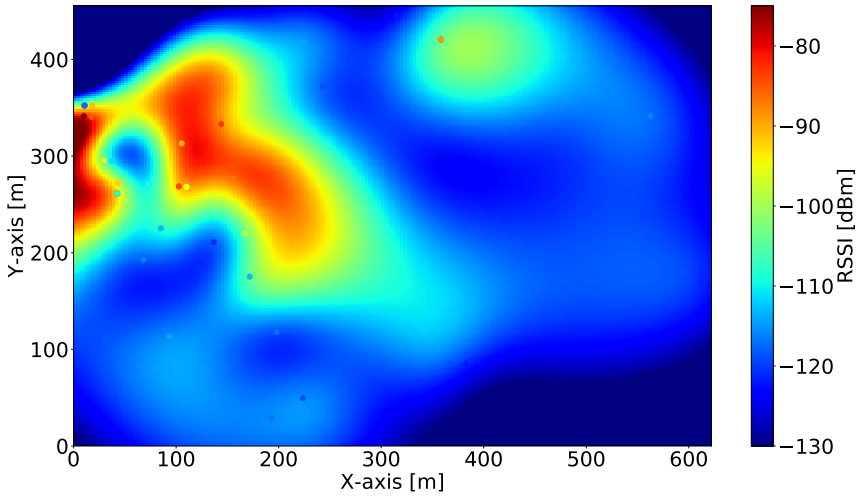
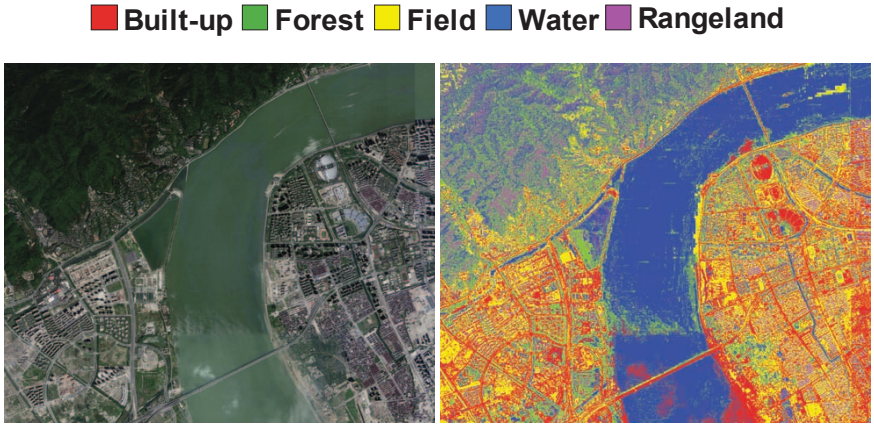


Figure 2.9 – Interpolated RSSI map at the Beaulieu campus by integrating 70 measured points (labeled by circles).



(a) Satellite image **(b) Land-cover map**

Figure 2.10 – An example of environment type type classification the in a 5 km × 5 km urban area of Hangzhou, China. The ground truth satellite image in RGB is converted to the corresponding environment type map by SateLoc [83].

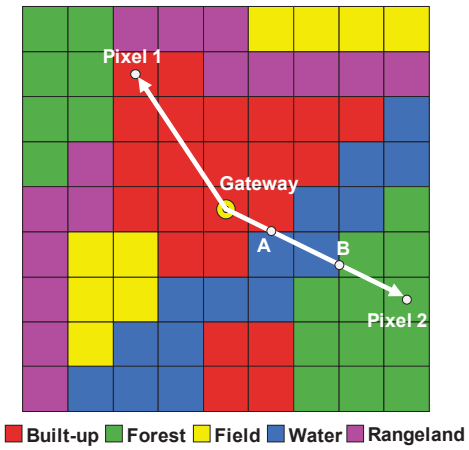


Figure 2.11 – Envision of a generated environment type map. Thus, LoRa node at Pixel 1 transmits its packets in the same environment type, i.e. built-up, while the packets need to propagate across multiple environment types node when transmitting from Pixel 2 [83].

2.4 Summary and Conclusion

This chapter summarizes the main signal features and the two major approaches for IoT localization. Although these approaches have different characteristics, they are complementary for localizing properly in different scenarios as follows:

- **Geometrical approaches:** For outdoor scenarios with open space, the geometrical methods take the lead as the path loss models can be explicitly modeled and parameterized. Moreover, the other signal measurements, for instance, ToA, TDoA or AoA values, are exposed to fewer error sources, such as reflection and multipath conditions, in such open environments.
- **Fingerprinting approaches:** In contrast, fingerprinting approaches are more suitable for indoor and urban scenarios with complex structures whose channel propagation properties are difficult to be modeled. With fingerprinting methods, these signal propagation phenomenons and multipath effects are exploited as additional information to enhance the localization accuracy. Subsequently, more potential is given to the fingerprinting methods for providing higher resolution and more details on complex localization scenarios.

For more accurate localization solutions, it is feasible to fuse different signal measurements [9, 84], or integrate the geometrical and fingerprinting methods [85, 86]. For example, [87] proposes a hybrid RSSI-AoA localization system for localizing a LoRa device with one gateway only. To estimate the node position, the distance between the gateway

and the node is calculated using a path loss model based on the RSSI value, after estimating the AoA of the received signal. This localization system is evaluated outdoor in a 55 m × 101 m playground in Kookmin University, Seoul, Korea. The results confirm the feasibility of estimating the end node position precisely using only one gateway in LoRaWAN.

However, these presented solutions are often sub-optimal when looking at what is needed for many IoT applications. Even the classical localization approaches should be optimized to meet the requirements for most IoT network operators, as detailed in the following chapters.

Chapter 3

Enhancement of the TDoA Technique

3.1 Introduction

The last decade has witnessed the development of the localization methods based on RSSI, AoA, ToA, TDoA and their multiple combinations [26], as detailed in the previous chapters. The gateways utilize these measurements to determine their relative position relations with the source for localization. In RSSI techniques, the existing relation between received power and the node position is exploited [9]. However, these techniques are sensitive to the channel environments and often require a good knowledge of the signal attenuation model [82]. In AoA techniques the angle from which the signal propagates is determined. The maintenance of an antenna array and precise calibration is always required, which results in extra expenses, moreover, their localization accuracy is sensitive to the distances between the node and the gateways [88]. ToA is one of the most accurate techniques available as it uses timestamps embedded in transmitted packets along with the received time to determine how far the packet had to travel to reach the destination [89]. Through the use of synchronized clocks, the signal propagation time between the transmitter and receiver can be determined. However, when using a ToA setup, devices in the network need synchronized clocks down to nanosecond scale in order to achieve a proper distance approximation, which requires additional hardware, thus increasing the cost of the system. Therefore, these three kinds of techniques are not so practical.

TDoA is similar to ToA but it is a more popular technique for localization as it does not require the transmitter to be synchronized with the receivers [4]. Only the gateways are required to have synchronized clocks to compute the differences between the timestamps of a transmission. After recalling from the previous chapters, these accurate timestamps are given at each gateway when receiving a packet by N gateways. These N gateways

are considered the anchor points with well-known locations at different distances to the end node, therefore, LoRa packet is received at different instances in time. Accurate time stamping is possible since each gateway has a GPS receiver for time synchronization. After acquiring the differences between the timestamps of a transmission, it is considered a multilateration problem that involves solving a set of hyperbolic functions, and therefore at least three gateways are needed to locate the node on the intersecting point of the hyperbolas. Based on that, the conventional geolocation for LoRa networks can be based on TDoA measurements, where hundreds to thousands of meters of accuracy can be achieved. In this chapter, the accuracy of the TDoA localization technique is improved by compensating for its shortcomings.

3.2 A Parametric TDoA Technique

3.2.1 Problem Statement and Contribution

In the practical situations of TDoA localization, the quality of each timestamp can vary largely from one gateway to another. This is due mostly to the time drift which occurs in the timestamping process and the effect of the multipath in Non-line-of-sight (NLoS) situations. These outliers timestamps may have an extremely deleterious effect on the final position accuracy with the classical TDoA solving techniques.

In this section, to tackle this problem, a novel parametric TDoA method is proposed that uses hyperbolic parameterization to localize the node on one of the hyperbolas instead of localizing it on the space far from the vicinity of the hyperbolas. Thus, the proposed approach is finding this location on a hyperbola at a point that has the minimum Euclidean distance to all the other hyperbolas. The accuracy and ease of use of this algorithm are evaluated through simulations. It is shown that the proposed approach effectively reduces localization errors in a wide range of situations covering different densities of gateways and noise models on ranging data.

3.2.2 Organization

The remainder of this section is organized as follows. Section 3.2.3 provides sufficient detail to allow implementation of the proposed algorithm. Section 3.2.4 provides guidance on the simulation model used to evaluate the performance of the algorithm on the

localization accuracy. The results of the simulation are then presented and commented in Section 3.2.5. Finally, Section 3.2.6 is dedicated for conclusions.

3.2.3 Parametric TDoA

Unlike the classical TDoA i.e., its system model is introduced in Section 2.3.1, in which the estimated position $\hat{\mathbf{p}}$ could be any point in the plane $\mathbb{R}^2$, the main concept of the parametric TDoA is obtaining the point $\hat{\mathbf{p}}_{h,i}$ to be placed on hyperbola i to be determined, by parameterizing the hyperbolic functions. This method is achieved using the following two steps:

1. **Minimizing the Euclidean distance matrix** by letting a set of points, each one lying in a different hyperbola among H hyperbolas to be expressed as:

$$\mathbf{P}_h(\mathbf{t}) = \Big[\mathbf{p}_{h,1}(t_1), \ldots, \mathbf{p}_{h,H}(t_H)\Big] = \begin{bmatrix} x_{h,1}(t_1), \ldots, x_{h,H}(t_H) \\ y_{h,1}(t_1), \ldots, y_{h,H}(t_H) \end{bmatrix}, \tag{3.1}$$

where $\mathbf{t} = \Big[t_1, \quad t_2, \quad \ldots \quad , t_H\Big]$ is the vector of the so called hyperbolic angles. Then, finding for every pair of gateways the distance between them as [90]:

$$\mathbf{d} = \Big[\|\mathbf{g}_{l(1)} - \mathbf{g}_{r(1)}\|_2, \ldots, \|\mathbf{g}_{l(H)} - \mathbf{g}_{r(H)}\|_2\Big] \tag{3.2}$$

and also the difference of distances vector $\mathbf{a}$ which is defined for each hyperbola as the timestamps difference at both gateways expressed in meter as:

$$\mathbf{a} = c \cdot \tau_{\mathbf{l},\mathbf{r}}, \tag{3.3}$$

while the distances from a focus to either asymptote i.e., the semi-minor axis, are defined as:

$$\mathbf{b} = \sqrt{\mathbf{d}^2 - \mathbf{a}^2}, \tag{3.4}$$

as well as, the exact centroid between each pair of gateways are calculated as:

$$\mathbf{q} = \frac{\mathbf{G_l} + \mathbf{G_r}}{2}. \tag{3.5}$$

Accordingly, all the H points on all the hyperbolas are initially acquired as:

$$\mathbf{P}_h(\mathbf{t}) = \mathbf{q} + \frac{\mathbf{a}}{2} \odot \cosh(\mathbf{t}) \odot \mathbf{U} + \frac{\mathbf{b}}{2} \odot \sinh(\mathbf{t}) \odot \mathbf{V} \tag{3.6}$$

with

$$\mathbf{U} = \frac{\mathbf{P}_h(t_\mathbf{r}) - \mathbf{P}_h(t_\mathbf{l})}{\mathbf{d}} = \left[\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_H\right] \tag{3.7}$$

and

$$\mathbf{V} = \left[\hat{\mathbf{z}} \times \hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{z}} \times \hat{\mathbf{u}}_H\right], \tag{3.8}$$

where $\odot$ and $\times$ are defined as the element wise matrix product and the vector cross product, respectively. Based on that, the obtained points $\mathbf{P}_h(\mathbf{t})$ have a deterministic Euclidean distance matrix defined as:

$$\mathbf{D}_h(\mathbf{t}) = \begin{bmatrix} 0 & d_{12}^2 & d_{13}^2 & \ldots & d_{1H}^2 \\ d_{21}^2 & 0 & d_{23}^2 & \ldots & d_{2H}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \ldots & d_{3H}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{H1}^2 & d_{H2}^2 & d_{H3}^2 & \ldots & 0 \end{bmatrix} \tag{3.9}$$

where $d_{ij}^2 = d_{ij}(t_i, t_j)^2 = \|\mathbf{p}_{h,i}(t_i) - \mathbf{p}_{h,j}(t_j)\|_2^2$ is the Euclidean distance between the hyperbola i and j. Here, the main target is to find the proper hyperbolic angles $\mathbf{t}_{min}$ which minimize the summation of the Euclidean distance matrix $\mathbf{D}_h(\mathbf{t})$, and this can be achieved using a solver as:

$$\mathbf{t}_{min} = \arg\min_{\mathbf{t}} \left(\sum_{i=1}^{H} \sum_{j=1}^{H} d_{ij}(t_i, t_j)^2\right). \tag{3.10}$$

2. **Electing the convenient point** from the set of estimated points $\mathbf{P}_h(\mathbf{t}_{min})$ is the final step i.e., one of the points among the small cross blue markers as shown in Figure 3.1. This is achieved by picking the index of the point which has the minimum Euclidean distance to all the other hyperbolas as:

$$i_{min} = \arg\min_{i} \left(\sum_{j=1}^{H} d_{ij}(t_{min,i}, t_{min,j})^2\right). \tag{3.11}$$

Hence, the final estimated location of the node is considered as:

$$\hat{\mathbf{p}}_{h,i_{min}} = \mathbf{p}_{h,i_{min}}(t_{min,i_{min}}), \tag{3.12}$$

which is the large cross blue marker in Figure 3.1, actually most of the cases closer to the true position than the classical TDoA estimation with the large cross green marker. The reason behind the proposed method is that it should be less sensitive to the outliers because of the constraint imposed on the points of belonging by construction to hyperbolas associated with the constraints.

3.2.4 Simulation Model

In this section, a brief overview of the simulator model is given in order to allow for a dynamic study of the system performance. Thus, the two main aspects shaping the simulation scenarios are the distributions of the gateway locations with respect to the node location and the choice of convenient uncertainty in the timestamps, to be closer to the real measurements as detailed in the following subsections.

Gateway locations

The gateway locations are produced by the Poisson disk distributions algorithm which has been introduced in [91]. This is considered as a fast 2-D blue noise sampler, easily implemented in arbitrary dimensions and it is guaranteed to take $O(M)$ time to generate M Poisson disk samples. To start the process, this algorithm takes as input the length *len* and width *wid* of the samples domain in $\mathbb{R}^2$, and the minimum distance ρ between the samples. First, it initializes a 2-D background grid for storing samples and accelerating the spatial searches. Then, it selects the initial sample randomly chosen uniformly from the domain and inserts it into a cell in the background grid. The cell size is picked to be bounded by $\frac{\rho}{\sqrt{2}}$, so that each grid cell will contain at most one sample. In the next iteration, the neighboring point is chosen uniformly from the spherical annulus between radius ρ and 2ρ around the previous sample. This linear algorithm is done recursively until all the M samples are generated. At this point, the node location is chosen randomly from the samples, while maintaining the other samples to be the gateway positions, as shown in Figure 3.1.

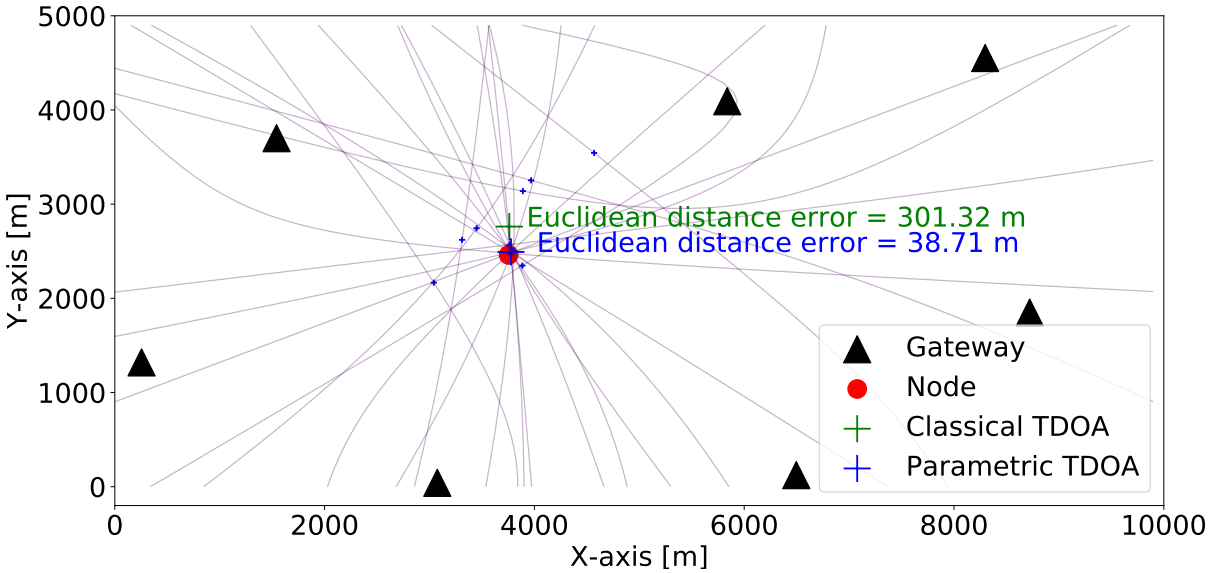


Figure 3.1 – A sample pattern from the algorithm ($\rho = 2.5\,\text{km}$, $wid = 10\,\text{km}$, $len = 5\,\text{km}$).

Timestamp perturbation model

From our knowledge of what is observed in LoRaWAN practical situations, it often happens that few gateways are suffering from much larger uncertainty than the others. To model this fact, a Gaussian mixture distribution is proposed. The first Gaussian is taking into account the measurement noise whereas the second is taking into consideration the larger oscillator drift perturbation. Hence, the timestamp is individually calculated after extending Equation 2.10 by modeling the uncertainty term u_n as:

$$\tau_n = \frac{d_n}{c} + u\,w_1 + (1-u)\,w_2 \tag{3.13}$$

with

$$\Pr(u=1) = p = 1 - \Pr(u=0), \tag{3.14}$$

where $w_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $w_2 \sim \mathcal{N}(0, \sigma_2^2)$ are the normal Gaussian distributions of the thermal noise and the gateway oscillator drift, respectively, by assuming zero means in both cases. Thus, the probability p is chosen equal to 0.8 taking into account the occurrence probability of the oscillator drift by 20%.

3.2.5 Simulation Results

In the previous section, the simulation framework is provided including the location distributions and the chosen noise model. In this section, the simulation results are presented, using the Euclidean distance error between the true position of the node and the estimated location, as a measure of performance for the localization methods. This is measured by Monte Carlo simulations after utilizing the proposed parametric TDoA technique and the classical TDoA using LS i.e., introduced in [66], for comparison. Moreover, the two main parameters which shape the simulation scenarios are the magnitude of the oscillator drift variance and the number of gateways, as detailed in the following subsections.

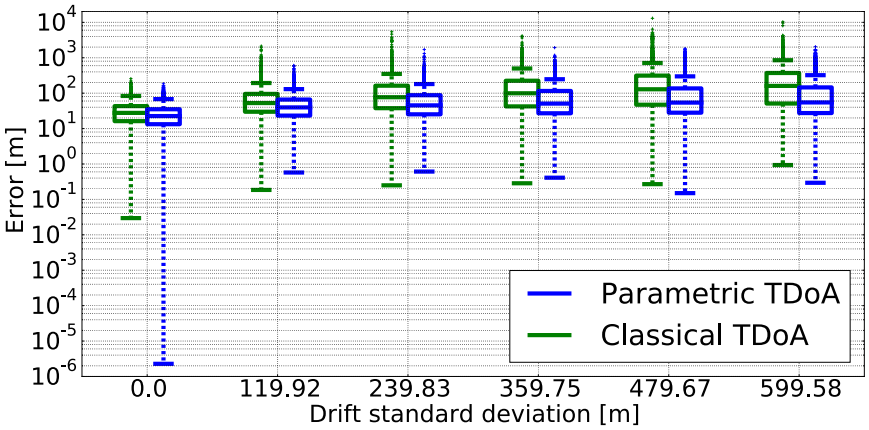
Impact of the oscillator drift variance

The robustness of the system against the outlier timestamps is checked by changing the drift standard deviation σ_2 in Equation 3.13 to be in the range from 0 µs to 2 µs ($\equiv$ 0 m to 600 m), while maintaining the thermal noise standard deviation σ_1 to be equal to 0.1 µs ($\equiv$ 30 m). For the gateway location distribution, the length *len* and width *wid* of the map are fixed to 5 km and 10 km, respectively, while the minimum distance ρ between the points is equal to 2.5 km. This distribution configuration usually gives a median number of gateways between 5 and 8.

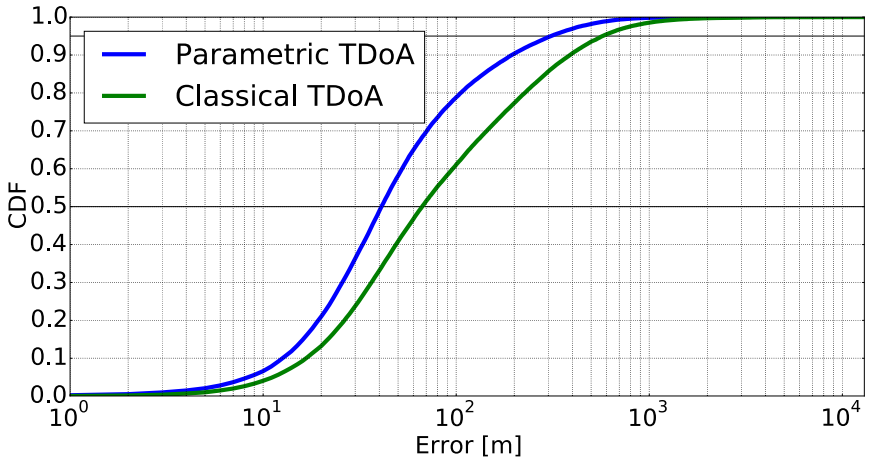
As shown in Figure 3.2a, the simulation result shows an obvious reduction in the localization error medians when using the parametric TDoA, especially for the large values of the drift. This indicates that the proposed method is more robust to the high drifts. Accordingly as shown in Figure 3.2b, the CDF curves obtained for all the drift values preserve the same performance rank over the whole simulations with 50% of the error values less than 40 m and 65 m, using the proposed parametric and classical TDoA, respectively.

Impact of the number of gateways

The localization performance assessment for various number of gateways is studied by utilizing the minimum distance ρ between the gateways to be in the range from 1.6 km to 4 km, while fixing the length *len* and width *wid* of the map to be 5 km and 10 km, respectively. Moreover in all the results, the drift standard deviation σ_2 and the thermal noise standard deviation σ_1 are assumed to be equal to 1.2 µs ($\equiv$ 360 m) and 0.1 µs ($\equiv$ 30 m),



(a) Errors at each value of the drift standard deviation.



(b) The whole simulation results of σ_2 ranging from 0 m to 600 m.

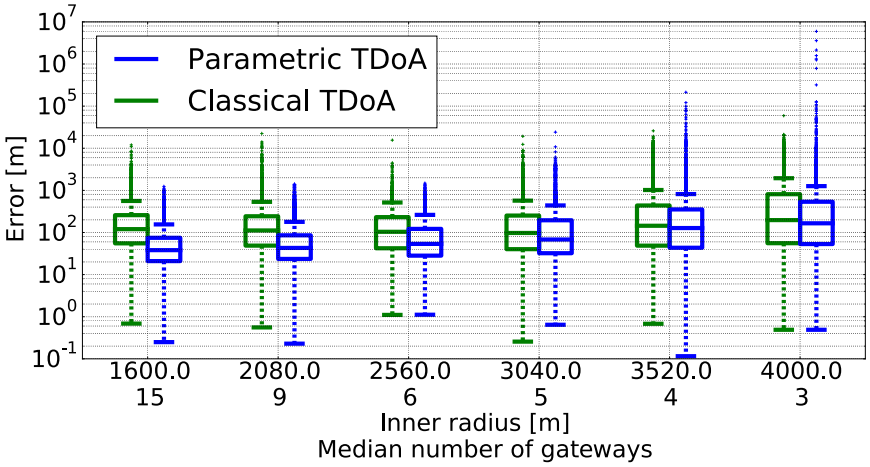
Figure 3.2 – Euclidean distance error and CDF while changing drift standard deviation σ_2.

respectively.

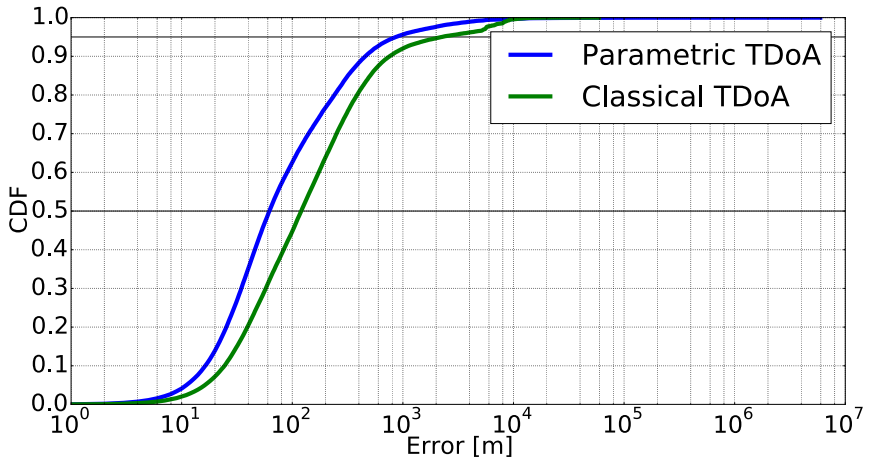
As shown in Figure 3.3a, it is clear that all the medians of the parametric method are drastically the lowest values for all the network densities. While by observing the extreme error values, the main parameter in achieving the performance promised by the proposed algorithm without these outlier values, is the large number of gateways (>4) i.e., a realistic value in the near future. Thus, it is clear in the lower network density that the number of hyperbolas decreases accordingly, resulting in an incorrect selection of the convenient point i.e., discussed in Section 3.2.1, from the set of estimated points $\mathbf{P}(\mathbf{t}_{min})$ and a less accurate localization. Only at these low number of gateways, there is a trade-off to find between the two methods, the parametric one being better regarding the median and worst regarding the outliers. Nevertheless as shown in Figure 3.3b, the CDF curves obtained for all the inner radius values still confirm the prevalence of the proposed method over most of the simulations with 95% of the error values less than 840 m and 2200 m, using the proposed parametric and classical TDoA, respectively.

3.2.6 Conclusion

In this section, a novel TDoA parametric method for IoT localization is presented. The technique improves the accuracy of localization in the presence of outliers, even when high levels of drift in gateways can hardly be avoided. The principle is based on the choice of the node position to be located on one of the hyperbolas by bringing the optimization into the parametric space of the hyperbolic constraint sets. For performance assessment of this technique, a simulator has been developed that uses a Poisson distribution approach for determining the location of gateways and nodes based on configurations chosen to be as close as possible to the realistic situations encountered in the LoRaWAN context. Moreover, a noise model is proposed to emulate the proper disturbance in the timestamp values by considering the drift variance. Simulation results prove the high performance of our parametric method over a wide range of drift standard deviations and network densities. In some configurations the achieved performance using the proposed technique, almost matches the performance in non-drifted timestamp scenarios. This competes particularly well regarding localization accuracy with more traditional approaches. This is going to become increasingly useful when situations of localization with a greater number of gateways will be more frequent in the future.



(a) Errors at each number of gateways.



(b) The whole simulation results of ρ ranging from 1.6 km to 4 km.

Figure 3.3 – Euclidean distance error and CDF while changing inner radius ρ.

For future work, parametric TDoA can be used to detect outliers as well as an initial guess point for more sophisticated localization approaches that integrate other a priori information.

3.3 A Pre-processing Algorithm Utilizing a Paired CRLB for TDoA

3.3.1 Problem Statement and Contribution

In the practical situations of TDoA localization, not all pairs of gateways are equivalent concerning the amount of extra information they bring to the estimation problem. Using the entire combinations as the classical TDoA solving techniques may have an extremely deleterious effect on the final position accuracy. Thus, it could be advantageous to define some criteria to select the best combination of gateways or equivalently the best hyperbolas to resolve the estimation problem.

In this section, to tackle this issue, a novel preprocessing method is proposed by utilizing a paired CRLB as an instrumental tool to detect the mostly perturbated TDoA values. This capability is due to the relative position of the node with respect to each pair of gateway locations which affect the paired CRLB values specifically. Accordingly, the perturbation probability in each TDoA value can vary largely. In the vicinity of the guessed node location, the proposed approach is detecting the best TDoA values, which have the lowest paired CRLB values specifically, based on a robust thresholding method. The accuracy and ease of use of this pre-processing algorithm are evaluated through simulations. It is shown that the proposed approach effectively reduces localization errors in a wide range of situations covering different densities of gateways and measurement noise variances.

3.3.2 Organization

The remainder of this section is organized as follows. Section 3.3.3 presents instructions to allow implementation of the proposed algorithm. Section 3.3.4 provides guidance on the simulation model used to evaluate the performance of the algorithm on the localization accuracy. The results of the simulation are then presented and commented in Section 3.3.5.

Finally, Section 3.3.6 is dedicated for conclusions.

3.3.3 Paired CRLB pre-processing

The CRLB is a lower bound on the covariance of any unbiased estimation algorithm based on the measurement [92]. This is calculated for the whole TDoA combinations from the inverse of the Fisher Information Matrix (FIM) $\mathbf{J}$ as:

$$E[(\hat{\mathbf{p}} - \mathbf{p_t})(\hat{\mathbf{p}} - \mathbf{p_t})^T] \geq \mathbf{J}^{-1}, \tag{3.15}$$

where $E[\cdot]$ determines the expectation value and $\mathbf{p}_t$ is the true node position.

As mentioned before, when dealing with a rather large number of gateway combinations, not all relative configurations with respect to the unknown node position are equivalent. Therefore, the proposed method invokes the paired CRLB information applied on each combination of 2 gateways to detect these most suitable hyperbolas associated with the best configurations. Thus, the node is localized after removing the combinations which have the highest paired CRLB values at an initial guess location $\mathbf{p}_o = [x_o, y_o]^T$ rather than, using the whole combinations including the ones whose configurations are less favorable and have been observed to introduce large errors. This algorithm is achieved using the following two steps:

1. **Estimating the paired CRLB values** is done separately for each combination by assuming a constant standard deviation σ for all the gateways. Thus, the covariance matrix at each pair of gateways is reduced to the scalar value $R = 2\sigma^2$. Accordingly, the Jacobian of the measurement function for a given pair of gateways i evaluated at position $\mathbf{p}_o$ is calculated as:

$$\begin{aligned}
\mathbf{f}_i^o &= \left.\frac{\partial h_{l(i),r(i)}}{\partial \mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}_o} && (3.16)\\
&= \left[\left.\frac{\partial h_{l(i),r(i)}}{\partial x}\right|_{\mathbf{p}=\mathbf{p}_o}, \left.\frac{\partial h_{l(i),r(i)}}{\partial y}\right|_{\mathbf{p}=\mathbf{p}_o}\right] && (3.17)\\
&= \left[\frac{x_o - x_{l(i)}}{d^o_{l(i)}} - \frac{x_o - x_{r(i)}}{d^o_{r(i)}}, \frac{y_o - y_{l(i)}}{d^o_{l(i)}} - \frac{y_o - y_{r(i)}}{d^o_{r(i)}}\right], && (3.18)
\end{aligned}$$

with $d^o_{l(i)} = \|\mathbf{g}_{l(i)} - \mathbf{p}_o\|_2$ and $d^o_{r(i)} = \|\mathbf{g}_{r(i)} - \mathbf{p}_o\|_2$.

Hence, the computation of the Fisher Information Matrix follows as:

$$\mathbf{J}_i^o = \frac{1}{2\sigma^2}\mathbf{f}_i^{oT}\mathbf{f}_i^o, \tag{3.19}$$

thus, the paired CRLB value is calculated from the inverse of the FIM $\mathbf{J}_i^o$, then, calculating the square root of its trace as:

$$c_i^o = \sqrt{tr(\mathbf{J}_i^{o-1})}. \tag{3.20}$$

2. **Determining the outliers** from all the determined paired CRLB values

$$\mathbf{C}^o = \left[c_1^o, \ldots, c_H^o\right] \tag{3.21}$$

, which are shown in Figure 3.4, by observing that the high peak at any combination index is considered as an outlier combination, while the low peaks are the proper combinations to be utilized for the localization. Thus, several threshold methods have been investigated, which all try to separate the outliers from the regular paired CRLB values without other a priori knowledge. These investigated thresholding techniques are mainly based on analyzing the variance of the paired CRLB values since it is observed that the regular values have amplitudes close to one another. Subsequently, all these methods assume that the combinations have two classes: ordinary paired CRLB value class and outlier class which are below and above the shown threshold in Figure 3.4, respectively.

In this section, the proposed thresholding technique is based on Median Absolute Deviation (MAD) which is a robust measure of how spread out a set of data is [93]. Hence, the classical standard deviation estimator could also measure the spread, but it is more affected by the extremely high or extremely low values and non-normality due to the presence of outliers. Accordingly, this threshold is computed as:

$$\gamma = \text{median}(\mathbf{C}^o) + 3 \cdot \hat{\sigma}_{MAD} \tag{3.22}$$

with

$$\hat{\sigma}_{MAD} = b \cdot \text{median}(|\mathbf{C}^o - \text{median}(\mathbf{C}^o)|), \tag{3.23}$$

where 3 has been heuristically chosen and the normalization factor b is set to 1.4826 in order to be consistent for the standard deviation at the normal Gaussian distribution i.e., usually the region of interest.

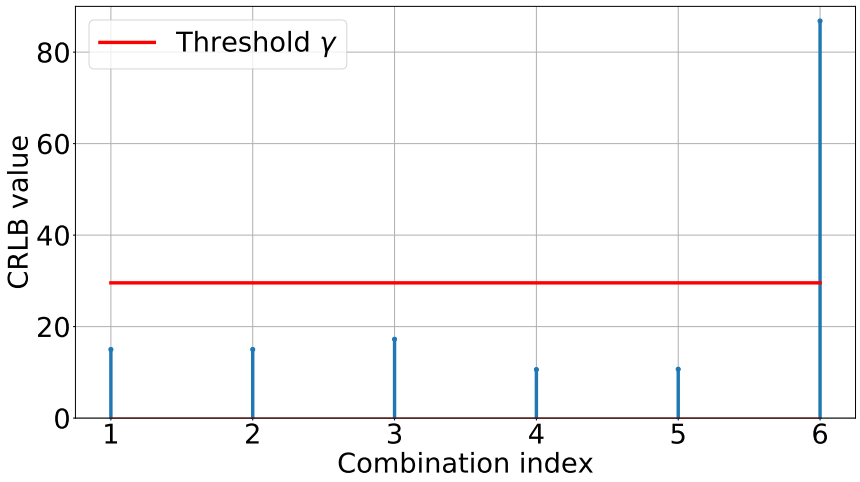


Figure 3.4 – The CRLB values of all the combinations ($H = 6$).

After labeling the outlier CRLB values by this threshold technique, the node is localized without these combinations, as shown in Figure 3.5. After removing the outlier combination based on its poor Geometric Dilution of Precision (GDoP), such a strategy reduces the error variance in the final estimated position.

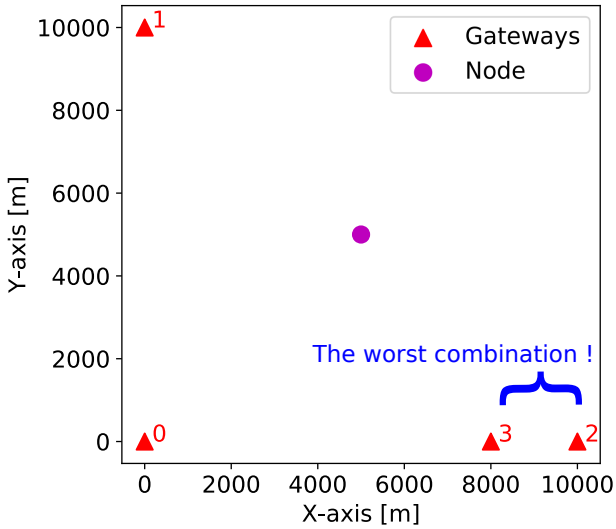


Figure 3.5 – A typical geometrical configuration of gateways with an outlier combination.

3.3.4 Simulation Model

Two main aspects shape the simulator model which allows for a dynamic study of the system performance. These are the distributions of the gateway locations with respect to the node location and the choice of convenient uncertainty in the timestamps, to be closer to the real measurements as detailed in the following subsections.

Gateway locations

The gateway locations are produced by the Poisson disk distributions algorithm which has been introduced in Section 3.2.4. After generating M samples, the node location is chosen randomly from the samples, while maintaining N extracted samples to be the gateway positions. Notice that generating the node location out of the dense M samples set forbids the node location to be less than ρ meter from any gateway.

Timestamp perturbation model

For the sake of simplicity, only the thermal noise using a normal Gaussian distribution is considered. Thus, the timestamp is individually calculated as in Equation 2.10 while assuming a zero mean of the uncertainty term $u_n \sim \mathcal{N}(0, \sigma^2)$.

3.3.5 Simulation Results

In the previous section, the simulation framework is provided including the location distributions and the chosen noise model. In this section, the simulation results are presented, using the Euclidean distance error between the true position of the node and the estimated location, as a measure of performance for the localization methods. This is measured by Monte Carlo simulations after utilizing the parametric TDoA technique i.e., introduced in [94], with and without paired CRLB pre-processing for comparison. Moreover, the two main parameters which shape the simulation scenarios are the magnitude of the thermal noise variance and the number of gateways, as detailed in the following subsections.

Impact of the noise variance

The robustness of the system against the outlier combinations is checked by changing the thermal noise standard deviation σ in Equation 2.10 to be in the range from 0.1 µs to

1 µs (≈ 30 m to 300 m), while maintaining the number of gateways to be $N = 7$ which are chosen randomly from a fixed location distribution. For this gateway locations configuration, the length *len* and width *wid* of the map are fixed to 5 km and 10 km, respectively, while the minimum distance ρ between the M samples is fixed to 100 m.

As shown in Figure 3.6a, the simulation result shows a noticeable reduction in the localization error medians when using the paired CRLB pre-processing algorithm. Accordingly as shown in Figure 3.6b, the CDF curves obtained for all the noise values preserve the same performance rank over the whole simulations with 50% of the error values less than 200 m and 230 m, while turning the proposed algorithm ON and OFF, respectively. This indicates that the proposed method is robust to the high noise variances.

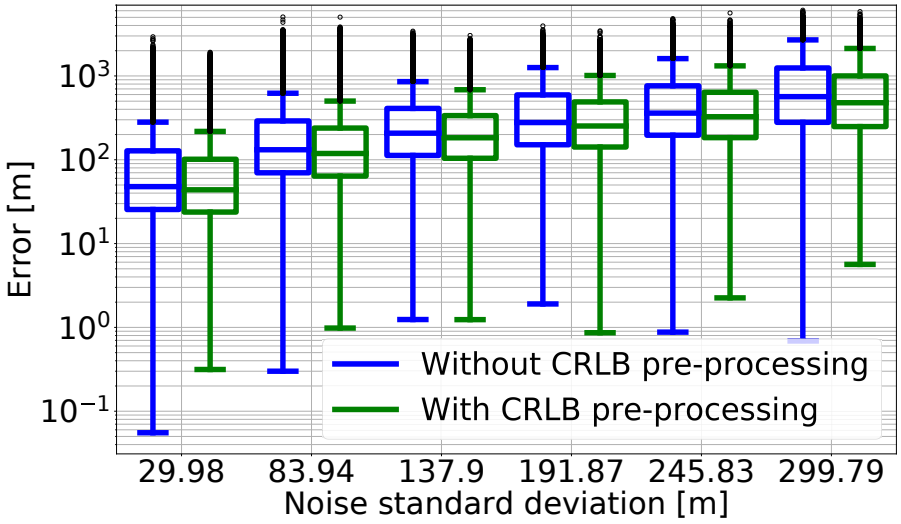
Impact of the number of gateways

The localization performance assessment for a various number of gateways is studied by utilizing the same gateway locations configuration i.e., introduced in the previous subsection, while choosing randomly from them a particular number of gateways N to be in the range from 3 to 15. Moreover in all the results, the thermal noise standard deviation σ is assumed to be equal to 0.1 µs (≈ 30 m).

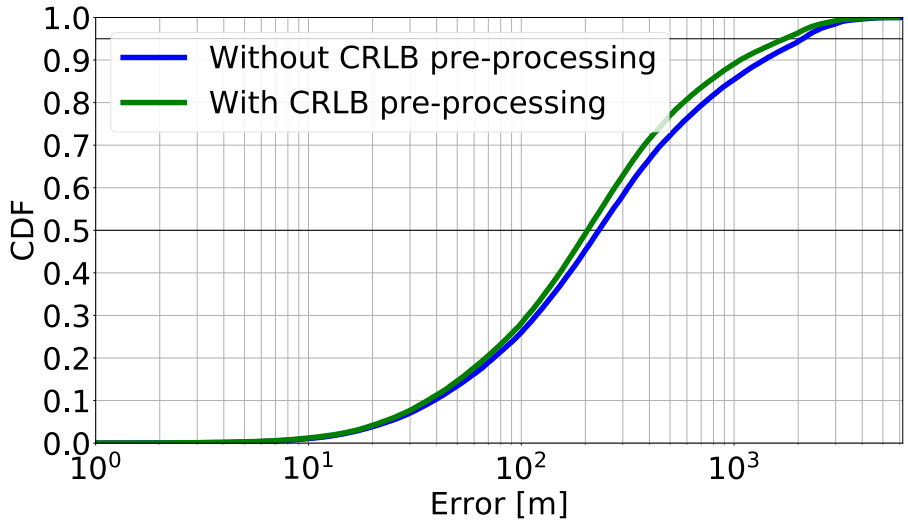
As shown in Figure 3.7a, it is clear that all the medians of the proposed method are drastically the lowest values for all the network densities, especially for a large number of gateways (> 3) i.e., a realistic value in the near future. Thus, it is clear in the lower network density that the number of hyperbolas decreases (≈ 3), accordingly, only one hyperbola might be dropped resulting in a tiny or almost no gain. Nevertheless as shown in Figure 3.7b, the CDF curves obtained for all number of gateways range still confirm the prevalence of the proposed method over most of the simulations with 95% of the error values less than 1600 m and 1940 m, while turning the proposed algorithm ON and OFF, respectively.

3.3.6 Conclusion

In this section, a novel TDoA pre-processing method for IoT localization is presented. This technique improves the accuracy of positioning in the presence of outlier hyperbolas by choosing only the best hyperbolic constraint sets for estimating the node location

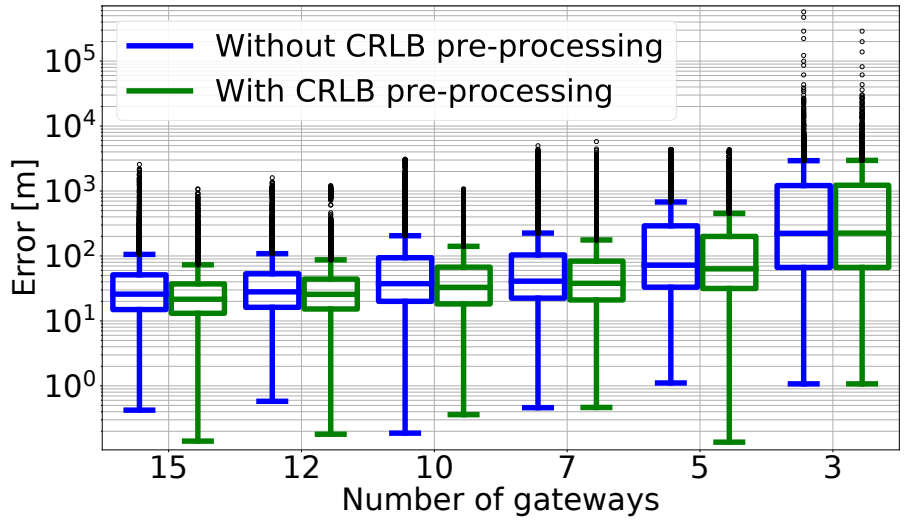


(a) Errors at each value of the noise standard deviation.

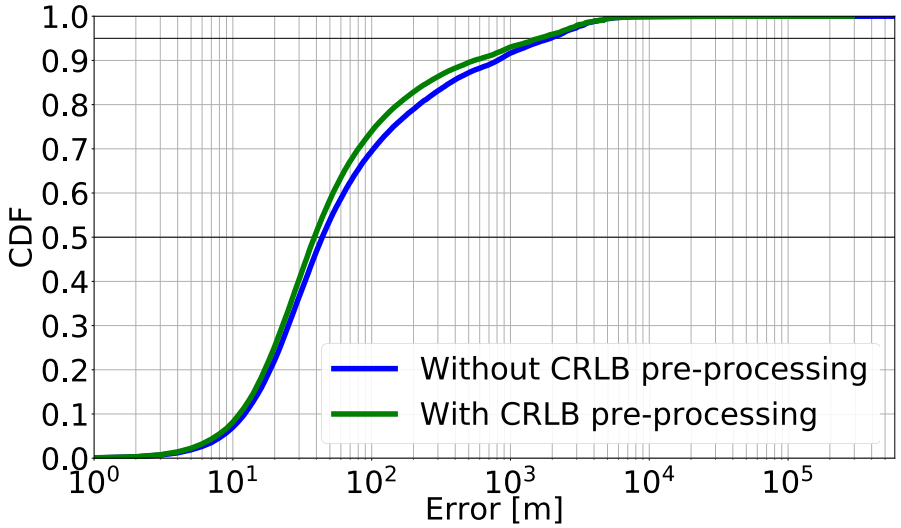


(b) The whole simulation results of σ ranging from 30 m to 300 m.

Figure 3.6 – Euclidean distance error and CDF while changing noise standard deviation σ.



(a) Errors at each number of gateways N.



(b) The whole simulation results of N ranging from 3 to 15.

Figure 3.7 – Euclidean distance error and CDF while changing the number of gateways N.

based on their low paired CRLB values. Thus, the principle is based on dropping the combinations, whose paired CRLB values are high in the proximity of the node location, using a robust thresholding technique. For performance assessment of this algorithm, a simulator has been developed. It uses a Poisson distribution approach parameterized for setting the location of gateways and nodes as close as possible to realistic situations. The used simulator also emulates the proper disturbance in the timestamp values. Simulation results demonstrate the high performance of our pre-processing algorithm over a wide range of noise standard deviations and network densities.

For future work, an optimum technique for outliers detection could be investigated rather than using a thresholding method.

3.4 A Pre-processing Algorithm for Outlier Timestamp Detection

3.4.1 Problem Statement and Contribution

In the practical situations of TDoA localization, the accuracy of each timestamp can vary largely from one gateway to another. This time drift in the timestamping process is due mostly to a large thermal noise or any imperfection that could accidentally happen in the process. Using the entire timestamps degrade the localization accuracy of the TDoA techniques.

In this section, to tackle this problem, a pre-processing method is proposed by utilizing an initial guess of the node position to detect the mostly perturbated timestamp values. This capability is due to the relative position of the node with respect to each hyperbola. First, the proposed approach is measuring the Euclidean distances from the guessed location to all the hyperbolas. Accordingly, the proposed approach detects the worst timestamps based on a robust thresholding method. These perturbated timestamps should be removed before using any TDoA localization technique, as detailed in the following subsections.

3.4.2 Outlier Timestamp Detection

The main concept of the proposed method is dropping the outlier timestamps which make some hyperbolas far away from the vicinity of the node location, as the bold hyperbolas in Figure 3.8. This method is achieved using the following two steps:

1. **Estimating the Euclidean distances** between the initial guess location $\mathbf{p}_o = [x_o, y_o]^T$ to all the hyperbolas is the first step. This initial guess location i.e., labeled by the large cross green markers in Figure 3.8, can be obtained by the parametric TDoA localization technique which is introduced in Section 3.2. Using Equation 3.10, the set of points $\mathbf{P}_h(\mathbf{t}_{min})$ which are lying on the H hyperbolas are acquired, as labeled by the small cross green markers in Figure 3.8. Then, the euclidean distances between each hyperbola and the initial guess of the node location are calculated as:

$$\mathbf{d}_h^o = \left[d_{h,1}^o, \ldots, d_{h,H}^o\right] = \left[\|\mathbf{p}_o - \mathbf{p}_{h,1}(t_{min,1})\|_2, \ldots, \|\mathbf{p}_o - \mathbf{p}_{h,H}(t_{min,H})\|_2\right]. \quad (3.24)$$

   As shown in Figure 3.9a, some of the euclidean distances have relatively extreme values than the other ones due to the inclusion of an outlier timestamp in the TDoA values which make these hyperbolas.

2. **Determining the outlier timestamps** is done by calculating first the average euclidean distances corresponding to each gateway index n as:

$$\mathbf{d}_a^o = \left[d_{a,1}^o, \ldots, d_{a,n}^o, \ldots, d_{a,N}^o\right] \quad (3.25)$$

   with

$$d_{a,n}^o = \frac{1}{\sum_{i=1}^{H} \mathbb{1}_{l(i)|r(i)=n}} \sum_{i=1}^{H} d_{h,i}^o \mathbb{1}_{l(i)|r(i)=n} \; \forall n. \quad (3.26)$$

   As shown in Figure 3.9b, the estimated average euclidean distances $\mathbf{d}_a^o$ have a high peak at gateway index 4 which is considered as an outlier, while the low peaks correspond to the proper timestamps to be utilized for the localization. As previously discussed in Section 3.3.3, several thresholding methods have been investigated, which all try to separate the outliers from the average euclidean distances without

other a priori knowledge. These investigated thresholding techniques are mainly based on analyzing the variance of the average euclidean distances since it is observed that the regular values have amplitudes close to one another. Consequently, all these methods assume that the average euclidean distances have two classes: ordinary average euclidean distances class and outlier class which are below and above the shown threshold in Figure 3.9b, respectively. Again, a thresholding technique based on MAD is proposed. Accordingly, this threshold is computed as:

$$\gamma = \text{median}(\mathbf{d}_a^o) + 3 \cdot \hat{\sigma}_{MAD} \tag{3.27}$$

with

$$\hat{\sigma}_{MAD} = b \cdot \text{median}(|\mathbf{d}_a^o - \text{median}(\mathbf{d}_a^o)|), \tag{3.28}$$

where 3 has been heuristically chosen and the normalization factor b is set to 1.4826 in order to be consistent for the standard deviation at the normal Gaussian distribution i.e., usually the region of interest.

After labeling the indices of the outlier timestamps by this thresholding technique, the node is localized without these timestamps. As shown in Figure 3.8, the euclidean distance error of the new estimated position i.e., labeled by a large cross blue marker, is reduced after dropping the outlier timestamps in comparison to the scenario with the large cross green marker before dropping the outlier timestamps. Such a strategy increases the accuracy of the estimated position as the estimation process is not attracted to the outlier hyperbolas i.e., relatively far away from the node position, after dropping the outlier timestamps.

3.5 Summary and Conclusion

This chapter introduces the proposed strategies and methods to improve the localization accuracy of TDoA localization. These proposed approaches could be used as independent or complementary methods to obtain a more precise localization. For complementary usage, the proposed approaches are recommended to be utilized sequentially in the following order:

1. **An outlier timestamp detection algorithm** is a proposed preprocessing algorithm to drop out the perturbated timestamps before localizing using any TDoA

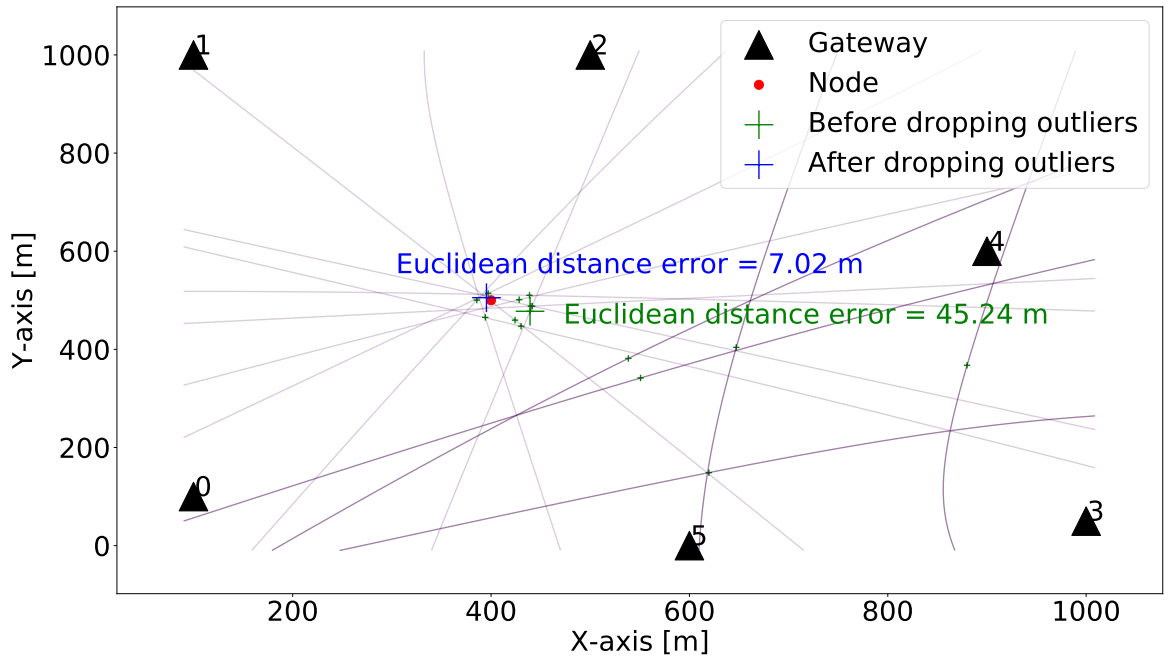
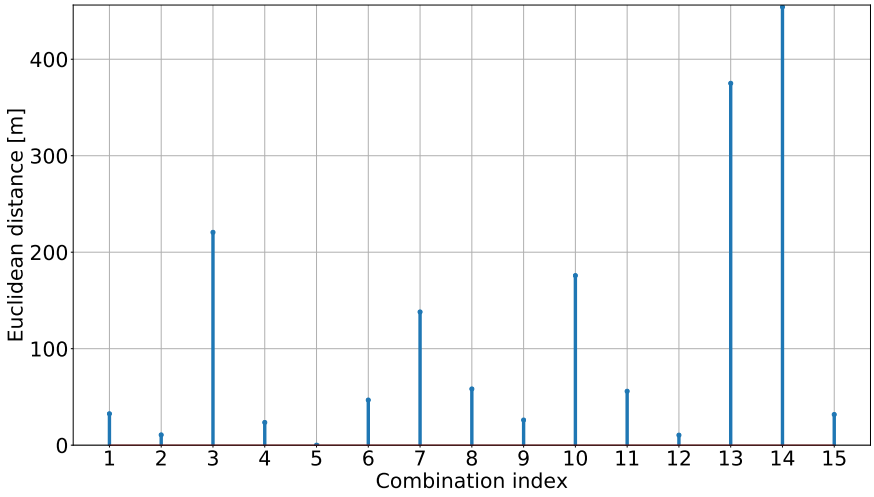


Figure 3.8 – Improvement of the localization accuracy after dropping the outlier timestamps.

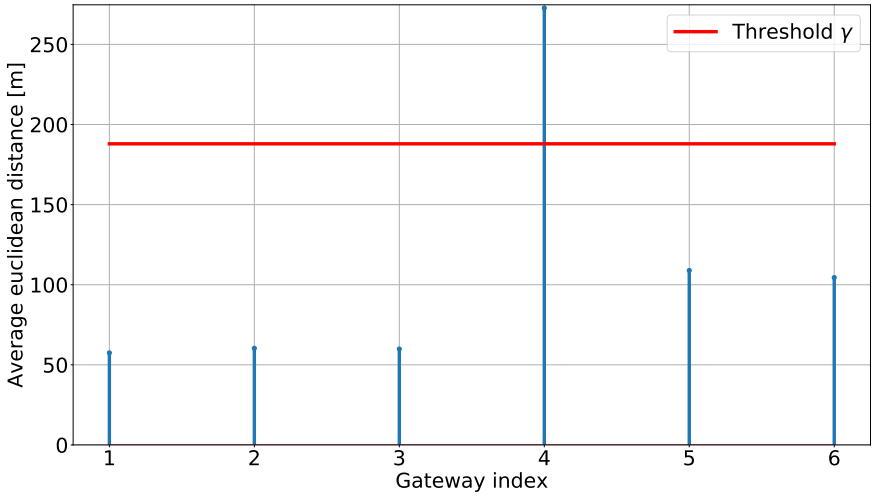
technique. An initial guess of the node location and the euclidean distances from it to each hyperbola are the only requirement for this approach. Then, the perturbated timestamps are detected from examining the most perturbed hyperbolas whose locations are far away from the vicinity of the node location. Dropping these outlier timestamps is proven to increase the localization accuracy by preventing any unreliable sensitivity from them.

2. **An outlier TDoA detection algorithm** is another proposed preprocessing method to detect the outlier TDoA values by utilizing a paired CRLB as an instrumental tool. From the vicinity of an initial guess of the node location, the TDoA combinations, whose paired CRLB are low, are dropped based on a thresholding technique. Using an implemented simulator, this proposed method is proven to decrease the localization error by localizing with the best TDoA values whose perturbation probabilities are less.

3. **A parametric TDoA technique** is a proposed approach to localize the node on one of the hyperbolas, contrary to the classical TDoA technique which localizes very often on the space far away from the vicinity of node location. The simulation



(a) Euclidean distances from the initial guess of the node location to all the hyperbolas ($H = 15$).



(b) Average euclidean distances corresponding to each gateway timestamp.

Figure 3.9 – Detecting the outlier timestamps.

results show that this proposed parametric method outperforms the classical one over different timestamps drift variances and number of gateways. Based on that, this proposed method is asserted to be less sensitive to any perturbation in the timestamps or the TDoA values.

For future work, the presented TDoA techniques can be complemented with classical machine learning techniques for merging other radio observations such as RSSI, AoA estimates or digital elevation model of the propagation environment.

Chapter 4

Feasibility of Utilizing the CSI for LoRa Localization

4.1 Introduction

Among all the localization technologies, wireless RSSI fingerprinting has proven as an effective positioning technique due to its simplicity and deployment practicability [82]. Fingerprinting based localization avoids hardware deployment cost and effort by relying on existing network infrastructure. It just relies on the received signal strength at each gateway to localize the node. However, RSSI-based fingerprinting localization methods may have poor positioning performance as RSSI always vary due to the large signal power fluctuations both in time and space, moreover, each measured RSSI value depends on the hardware accuracy as well as the system calibration for every measurement [49]. Besides, RSSI contains coarse information so as to not fully utilize the abundant channel information in each subcarrier. Therefore, some little work has been published whose aim is to take the average value of the whole CSI subcarrier amplitudes which is proven to be more temporally stable in different environments and helps maintain the performance over time in comparison with RSSI [95]. This chapter investigates the feasibility of using the CSI for localization in LoRaWAN, moreover, the advantage of utilizing ESP rather than RSSI in many prospective IoT applications is investigated.

4.2 Related Works

Recently, different characteristics of LoRa propagation are mentioned in many works. For example, the physical and data link layer performance are evaluated by field tests and simulations in [15]. While the coverage and channel attenuation of LoRa technology are evaluated for real environments in [96] and [97], path loss models are developed and compared with widely used empirical models based on empirical results in [98]. On the other

hand, the feasibility of utilizing the CSI for localization is studied in [99]. However, these previous works don't investigate either the spatial correlation of CSI or, the reciprocity between the uplink and downlink CSI which could be exploited in some security procedures in the LoRaWAN [100]. Moreover, these previous works do not inspect or utilize the ESP, i.e. introduced in [101].

4.3 A Robustness Comparison of Measured Narrowband CSI vs RSSI for IoT Localization

4.3.1 Contribution

In this section, it is favorable to show the possibility to leverage CSI for improving the performance of positioning by investigating the profit of using the entire subcarrier magnitudes without averaging or any reduction. Thus, an initial measurement campaign is done to compute the narrowband CSI of transmitted LoRa signals from different locations. To achieve significant localization gain, it is necessary for the individual channels from each different position to be uncorrelated with one another. Thus, the presented short time data indicate that channel slopes with even short separated distances will be quite stable and show weak intercorrelations between them. Furthermore, the CSI in each position after a while is more correlated with itself which can achieve a significant diversity gain in comparison with the mean amplitude of the CSI. This trial allows us also to derive recommendations for the use of diversity at the receiving site in short-range outdoor-to-outdoor transmission systems, asking questions like: "At what distance must vertically polarized antennas be placed such that intercorrelation is low and hence gateway diversity is potentially beneficial for localization?" or "Can also the LoRa frequency hopping add a diversity gain for a static transmitting node scenario?".

4.3.2 Organization

The remainder of this section is organized as follows. Section 4.3.3 presents the measurement overview and Section 4.3.4 provides sufficient detail of the proposed post-processing algorithm. The results of the channel correlation analysis are then presented and commented in Section 4.3.5. Finally, Section 4.3.6 is dedicated for conclusions.

4.3.3 System and Measurement Setup

The main concept of the proposed experiment is transmitting repeated up-chirps signal to sense the channels consecutively at the typical uplink frequency bands, i.e 9 channels with center frequency $f_k \in \{867.1, 867.3, 867.5, 867.7, 867.9, 868.1, 868.3, 868.5, 868.8\}$ MHz, and 125 kHz bandwidth. This is considered as a traditional channel sounder with a typical structure which has the Tx and Rx placed at two different locations, as shown in Figure 4.1. Thus, the Rx antenna is fixed on the roof of the university building, as shown in Figure 4.2a. While the Tx has a mobile structure with a laptop and a Universal Software Radio Peripheral (USRP), as it is described in Figure 4.2c. First, the Tx, i.e. located in specific positions within the area of the Campus Beaulieu in Rennes, should generate a signal by a laptop and transmit it using the USRP to sample the channel for a specific time interval. While the stationary Rx, whose antenna is located above the building, should receive the signal with its USRP at the same time interval of transmitting, as shown in Figure 4.2b. After the Rx picks up the signal, the desktop computer stores it to perform an essential post-processing algorithm on the received signal to mitigate the imperfections and obtain the channel transfer function, as detailed in the following subsection.

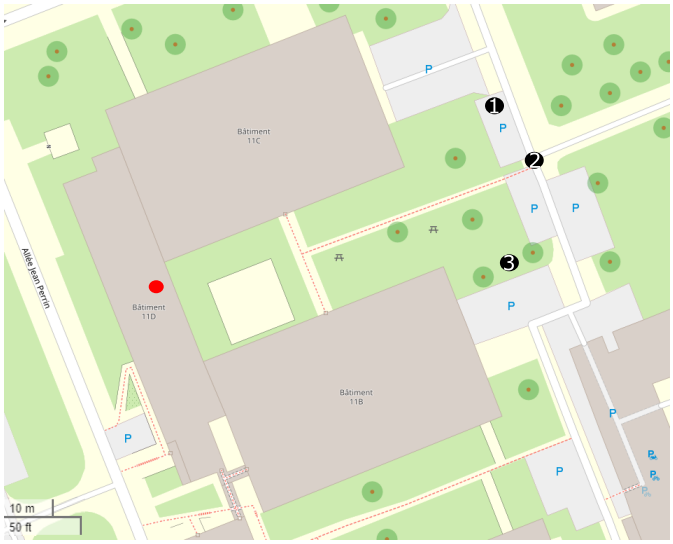
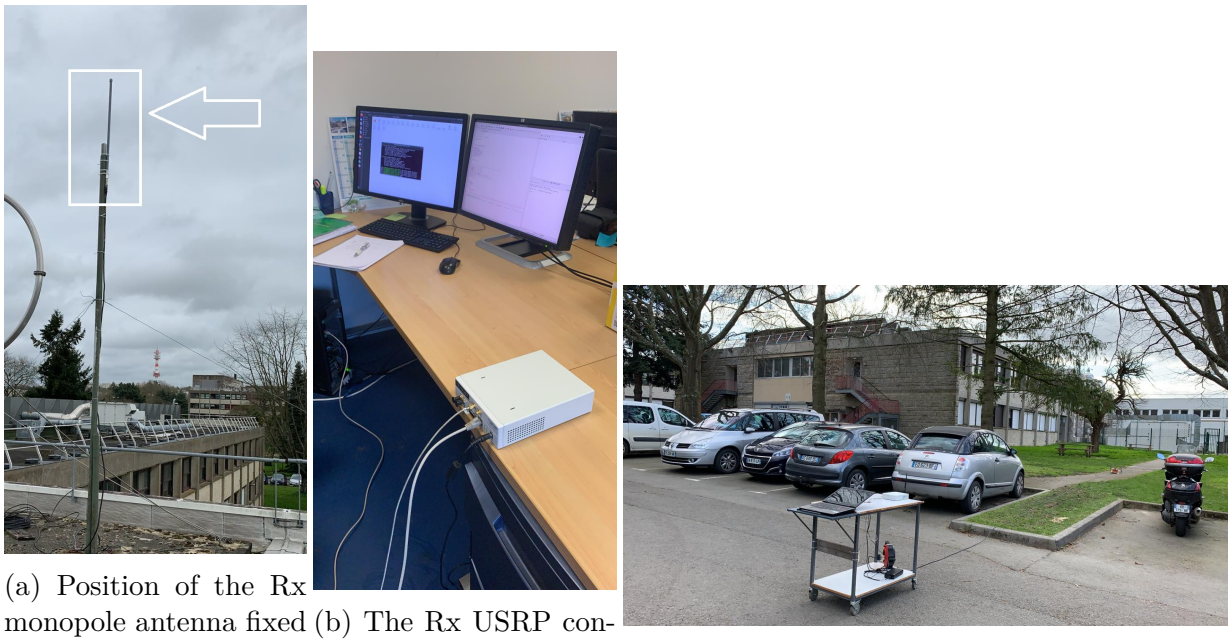


Figure 4.1 – The locations of the three measured points in the area of the Campus Beaulieu. Positions of Tx and the Rx are labeled by black, and red markers respectively. (©by OpenStreetMap Contributers)

Whereas the aforementioned emulated preamble LoRa signal is generated using python

such that $x[n]$ is expressed as depicted in equation 2.1, while the SF is chosen to be equal to 7 and $m = 0$ for the basic chirp symbol.



(a) Position of the Rx monopole antenna fixed on the roof of the IETR lab.

(b) The Rx USRP connected to the computer and the antenna cable.

(c) A trolley shelf with the different parts of the transmission equipment at Tx location 2.

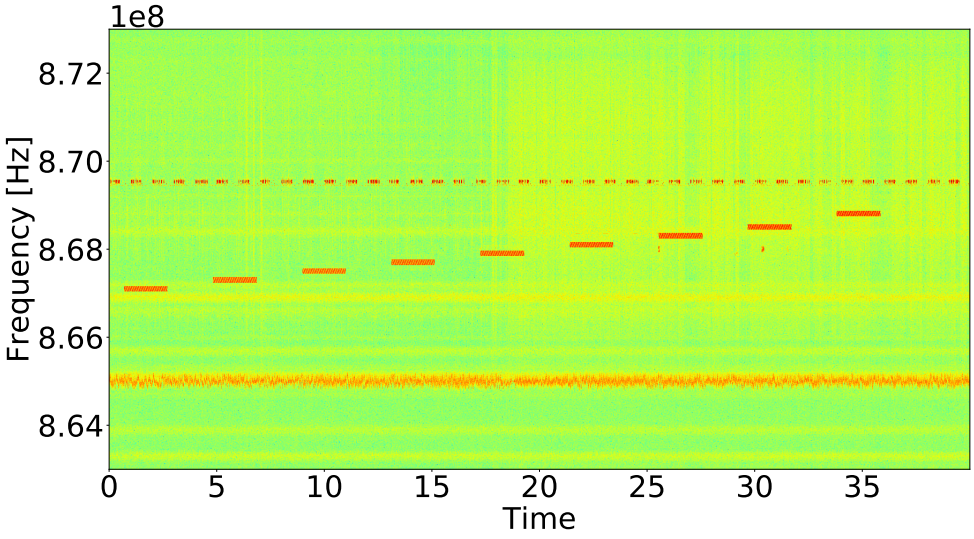
Figure 4.2 – Views from the Tx and the Rx sites.

4.3.4 Data Processing

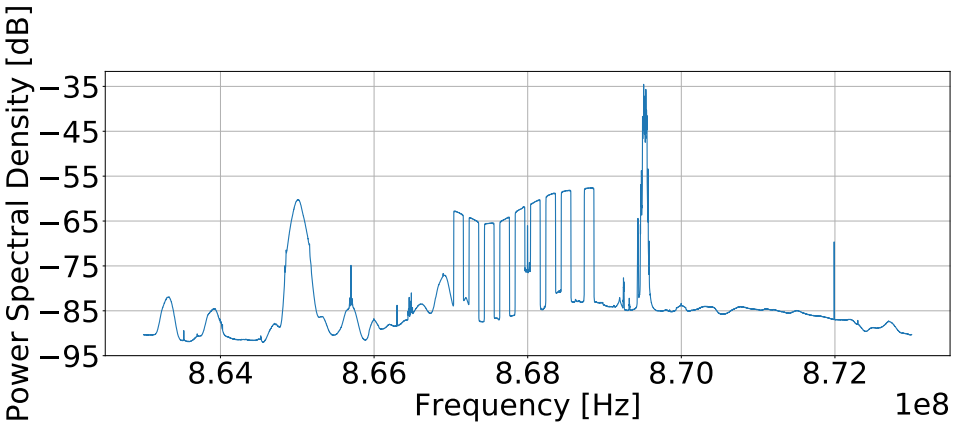
On the Rx side, the saved file is imported to be analysed for each center frequency f_k, as shown in Figure 4.3. Thus, the following main signal processing techniques are carried out with the same order.

Frequency synchronization

The frequency and time synchronization are applied at first, respectively. The objective of frequency synchronization is to establish the subcarrier orthogonality by correcting the phase as an initial step before applying any further processing [102]. Let define the Carrier Frequency Offset (CFO) by Δf_k as being the difference between the up and down conversion frequencies. This CFO results in a phase offset $\Delta\phi_k = 2\pi\frac{\Delta f_k}{f_s}$ between two samples with the same index in consecutive upchirps. The residual part of this offset can



(a) The spectrogram of the raw signal.



(b) The PSD of the raw signal.

Figure 4.3 – A typical 40 s of received data from location 2.

be estimated by taking the average across the received entire symbol $y_k[n]$ as:

$$\widehat{\Delta\phi_k}[l] = arg\ (\sum_{n=0}^{2^{SF}-1} y_k[l+n]y_k^*[l+n+2^{SF}]). \tag{4.1}$$

This method of detecting the CFO is described as the frequency acquisition algorithm and is utilized over the whole handled signal portion. As shown in Figure 4.4, the angles of the differential correlation function indicate that there are some phase deviations. Hence, the phase error is compensated and the angles of the corrected differential correlation function become concentrated around zero.

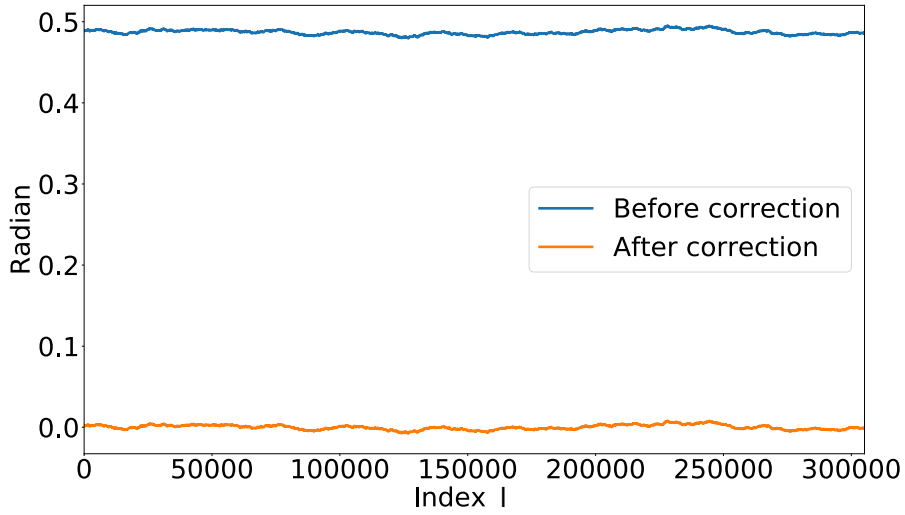


Figure 4.4 – The angles of differential correlation function $\widehat{\Delta\phi_k}[l]$ before and after correction of the whole portion of about 30 symbols.

Channel estimation

The up-chirps in the LoRa preamble, i.e. typically consist of eight symbols, are considered as a channel sounder. Therefore, the least squares estimate of the raw channel transfer function $\mathbf{H}_{raw}$ can be estimated for a simple division as[103]:

$$\mathbf{H}_{raw} = \frac{\mathbf{Y}}{\mathbf{X}}, \tag{4.2}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the training and the received symbol in the frequency domain respectively. Hence, the proposed denoising technique is applied on the raw channel transfer function $\mathbf{H}_{raw}$. This imposes doing an incipient step, i.e. removing the thermal noise from the subcarriers in the channel transfer function. Thus, the raw channel transfer function $\mathbf{H}_{raw}$ is low-passed in the frequency domain using Finite Impulse Response (FIR) filter to maintain the property of linear phase as:

$$\mathbf{H}_{filtered} = \mathbf{R}_h \cdot \mathbf{H}_{raw}, \tag{4.3}$$

where $\mathbf{R}_h$ is the autocorrelation matrix of the channel and the filtered channel $\mathbf{H}_{filtered}$ is also defined as $\mathbf{H}^l_{f_k}$. Where f_k and l are the center frequency value in MHz and the location number, respectively. As shown in Figure 4.5, one can observe that the obtained CSI is smooth without any noise at the three different locations. Moreover, it is obvious that the channels have a different magnitude of attenuation at the same subcarrier through the different locations.

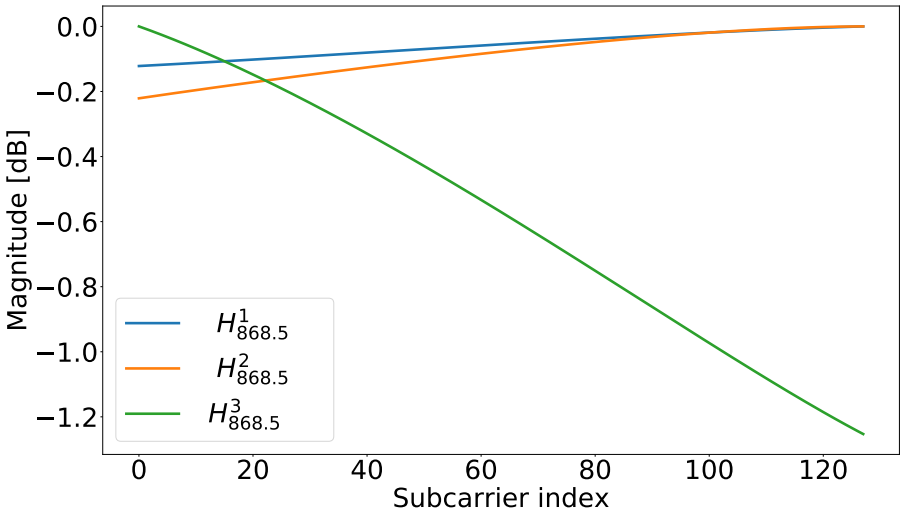


Figure 4.5 – The normalized CSI with 125 kHz bandwidth.

4.3.5 Correlation Analysis

To check the plausibility of utilizing the CSI for localization, the CSI spatial and temporal evolution at each location is analysed using a proposed method based on the CSI slope. Furthermore, this technique is also compared to the traditional method which

determines each position from its CSI average amplitude, as detailed in the following subsections.

CSI slope

Based on the LoRa narrow-band characteristic, i.e. with a bandwidth $W = 125\,\text{kHz}$ in Europe, a linear variation of the channel transfer function is assumed over frequency bandwidth centered on f_k. This condition is generally valid under the considered reality of such a very narrow bandwidth, thus, all the obtained channel transfer functions are almost flat, as shown in Figure 4.5. Nevertheless, it is clear that each CSI has a unique slope. In this section, the CSI slopes are estimated to analyse the CSI evolution from each location i as:

$$s^i_{f_k} = \frac{\mid \mathbf{H}^i_{f_k}[0] \mid - \mid \mathbf{H}^i_{f_k}[2^{SF}-1] \mid}{W}. \tag{4.4}$$

Then, the Normalized Slope Distance (NSD) between any two positions for the whole K uplink frequency bands is defined as:

$$S_{i,j} = \frac{1}{2K} \sum_{k=0}^{K-1} \mid \tilde{s}^i_{f_k} - \tilde{s}^j_{f_k} \mid \tag{4.5}$$

with

$$\tilde{s}^i_{f_k} = \frac{s^i_{f_k}}{S_{max}}, \tag{4.6}$$

where S_{max} is equivalent to the absolute value of the maximum observed slope in the measured dataset. This is chosen to be the scaling factor for the slope normalization. Consequently,

$$0 \leq S_{i,j} \leq 1. \tag{4.7}$$

The magnitudes of the aforementioned normalized slope $\tilde{s}^i_{f_k}$ are given in Figure 4.6a for each different location i and center frequency f_k. Thus, the full scale is normalized and lies in $[-1, 1]$, whereas every amplitude can vary on a several order of magnitude levels. Each frequency of the K uplink frequency bands is labeled with a specific color. In order to analyse the time variation of the CSI slope at each position, over a time duration τ of about 10 minutes, an arrow is drawn for every center frequency from the first time instant

to the second one. This measurement dataset demonstrates the slow alteration of the CSI slope with respect to time.

Moving to the NSD $S_{i,j}$ values in Figure 4.7a, one can observe the high level of NSD values in the three location combinations, which are $S_{1,2}$, $S_{1,3}$ and $S_{2,3}$. This indicates that CSI models from different positions are distinctly different. While the CSI for each location significantly underlines a low NSD value with itself after a τ time interval. This is obvious in $S_{1,1^\tau}$, $S_{2,2^\tau}$ and $S_{3,3^\tau}$ whose values are near zero as well as they are far away from the other NSD values, i.e. labeled by circles. So the environment of the propagation paths for each specific position is stable with only marginal modification during this duration.

CSI average magnitude

On the other hand, the channel evolution at each location is investigated by the mechanism, which depends on calculating the CSI mean magnitude. This method is considered as an alternative way rather than estimating the RSSI value. Thus, the CSI average amplitude is estimated for each location i as:

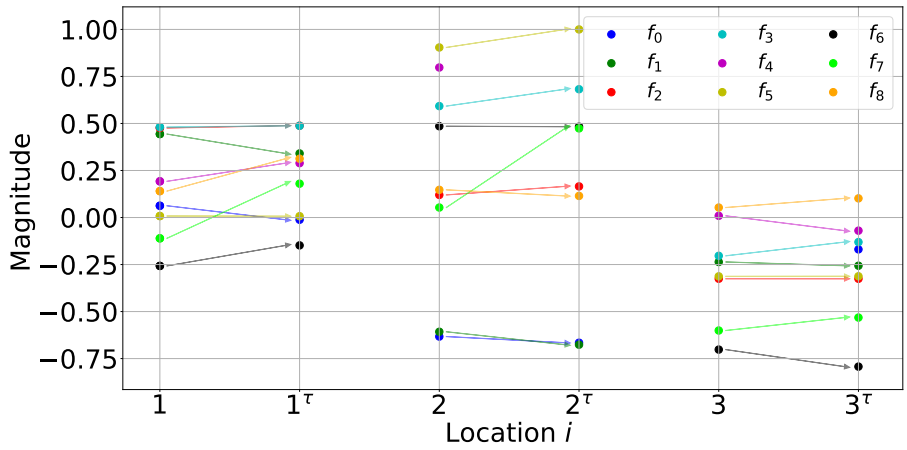
$$r^i_{f_k} = \frac{1}{2^{SF}} \sum_{n=0}^{2^{SF}-1} \mid \mathbf{H}^i_{f_k}[n] \mid, \tag{4.8}$$

hence, the Amplitude Distance (AD) between any two positions for the entire K uplink frequency bands is computed as:

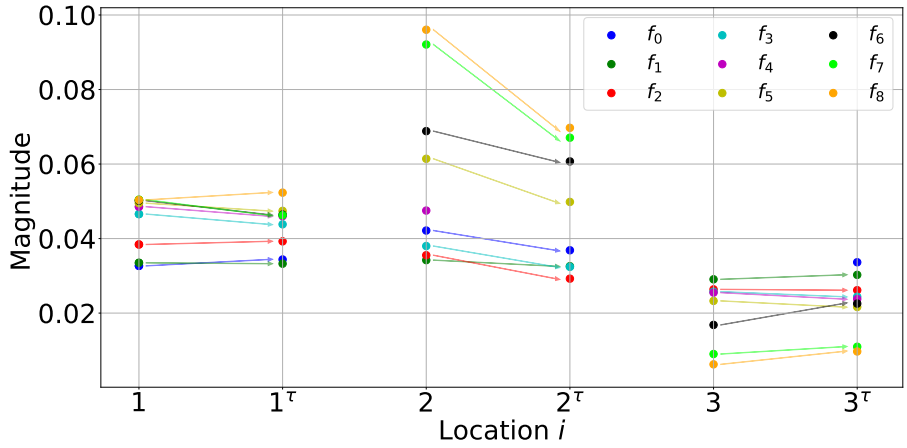
$$R_{i,j} = \frac{1}{K} \sum_{k=0}^{K-1} \mid r^i_{f_k} - r^j_{f_k} \mid . \tag{4.9}$$

For the whole evaluated locations and frequency bands, the measured values of the CSI average amplitude $r^i_{f_k}$ are given in Figure 4.6b. It is foremost supposed that the CSI slope variability from one sub-band to another is more informative than in the CSI average magnitude whose amplitudes at various bands are evaluating roughly at near levels. Moreover, it seems that the CSI average amplitude has more alternation in time than the CSI slope, which is particularly clear at location 2 when considering the shift between the values of 2 and 2^τ.

As depicted in Figure 4.7b, the AD $R_{i,j}$ values obtained for all the location combinations preserve proper values. However, the AD values, i.e. labeled by triangles, for each location with itself after the time interval τ are near to the AD values of different location combinations, i.e. labeled by circles. This could be observed explicitly in the high AD value of $R_{2,2^\tau}$. This confirms the proposed hypothesis which asserts that the CSI is more robust to the signal fluctuations than the RSSI, because, the CSI average magnitude values don't have the same temporal stability as the CSI slopes.

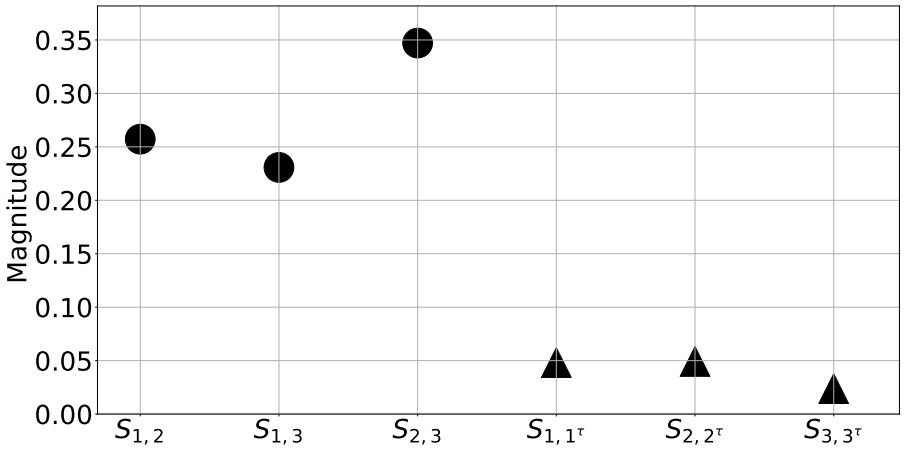


(a) The normalized slope $\tilde{s}^i_{f_k}$ values.

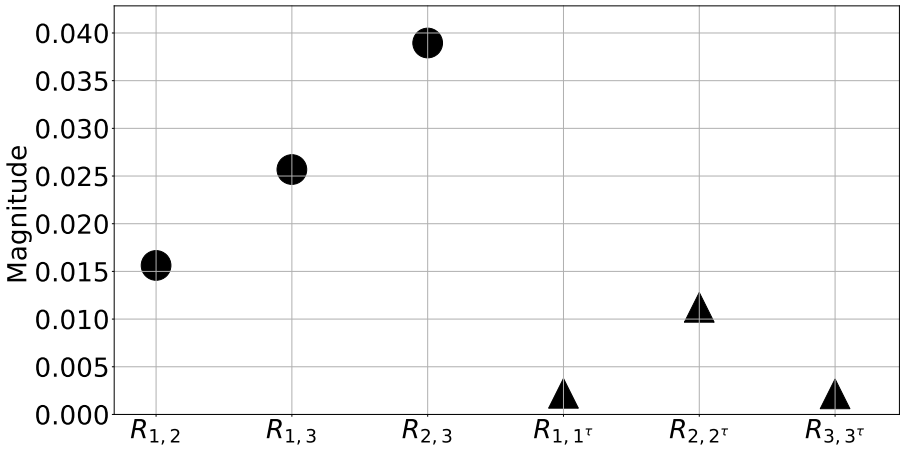


(b) The CSI average amplitude $r^i_{f_k}$ values.

Figure 4.6 – A comparison between the two families of the distinct observables.



(a) The NSD $S_{i,j}$ values.



(b) The AD $R_{i,j}$ values.

Figure 4.7 – The distance values.

4.3.6 Conclusion

Rather than using RSSI fingerprinting, in this section, the feasibility of utilizing the CSI for localization is presented. This hypothesis intends to improve the accuracy of positioning by utilizing the rich channel information in each subcarrier as well as the frequency hopping in the LoRa systems. To allow a dynamic study of this approach, an outdoor measurement campaign is carried out in the area of Beaulieu Campus in Rennes to estimate the CSI of transmitted LoRa signals from different locations. Hence, the individual channels from each different location have to be appropriately different with one another to achieve significant localization gain. This difference in the CSI for every two locations is done based on two aspects, i.e. CSI slope and its average amplitude. Thus, the presented data indicate that CSI slopes are more stable and robust to the signal imperfections than the CSI average amplitudes. This result demonstrates the high performance of the CSI-based fingerprinting for positioning than RSSI, as well as its temporal stability.

For future work, CSI-based fingerprinting could be more efficient by using more than one gateway i.e., a realistic value in the near future, to obtain more than one CSI for the instant singular center frequency. Moreover, it could be improved with classical machine learning techniques for merging other radio observables such as RSSI, AoA estimates or the propagation model.

4.4 Spatial Correlation of CSI in Real LoRa Measurement

4.4.1 Contribution

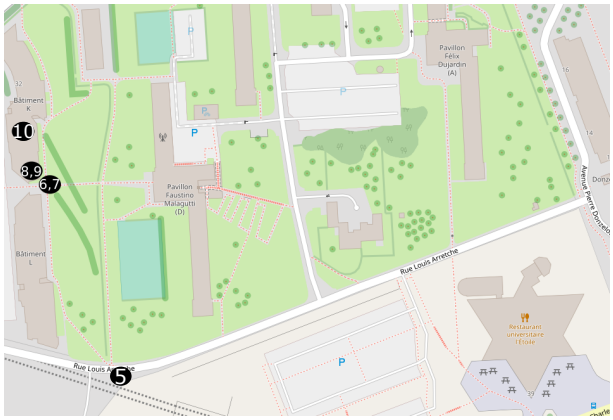
In this section, a real network deployment in the area of Beaulieu Campus in Rennes is employed to investigate the CSI at various locations. Hence, the ESP is used for the investigation to overcome the limitation of the RSSI at low SNR ($< 0\,\mathrm{dB}$) as proven. Thus, ESP values are average at each frequency band to estimate the absolute CSI of the uplink and downlink transmission. Consequently, the CSI reciprocity, as well as the spatial evolution of the CSI, are analyzed using the Pearson Correlation Coefficient (PCC).

4.4.2 Organization

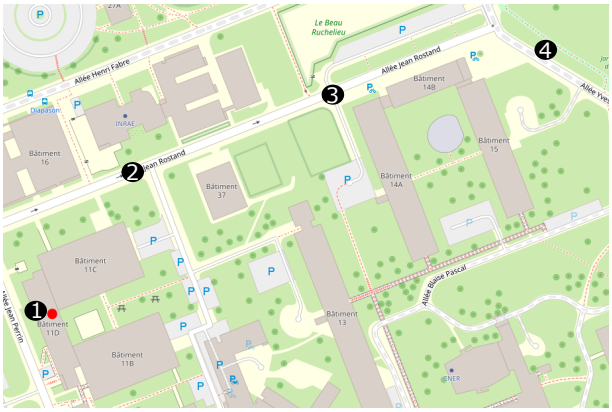
The remainder of this section is organized as follows. Section 4.4.3 presents the measurement overview. Section 4.4.4 provides sufficient detail about the derivation and obtaining of ESP values. Section 4.4.5 provides sufficient detail about the estimation of the CSI shape. The reciprocity between uplink and downlink CSI is then investigated in Section 4.4.6. Hence, Section 4.4.7 provides an evaluation of the spatial correlation between the CSI shapes. Finally, the work is concluded in Section 4.4.8.

4.4.3 Measurement Overview

The main aim of the proposed experiment is obtaining an averaged ESP value at each frequency band. This is done by placing an IoT node at various distances in the range of $\approx 5\,\mathrm{m}$ to $\approx 800\,\mathrm{m}$, as shown in Figure 4.8. While setting the LoRaWAN Configuration as depicted in Table 4.1. At the typical frequency bands for Europe, i.e. 8 channels with center frequency f_k, confirmed LoRa packets are transmitted sequentially with 125 kHz bandwidth W.



(a) The end node locations in the area of the Residence University Beaulieu. Location 8, 9 and 10 are indoor.



(b) The end node locations in the area of the Campus Beaulieu. Location 1 is indoor.

Figure 4.8 – Positions of the node and the gateway are labeled by black, and red markers respectively. (©by OpenStreetMap Contributers)

A Tektelic KONA Macro Gateway is used whose antenna is fixed on the roof of the university building [104], as shown in Figure 4.9a. The end node is implemented using a Pycom card, i.e. programmed in the MicroPython language, composed of an Expansion Board and a LoPy 4 module which can support LoRa wireless connectivity [105], as shown

Table 4.1 – LoRaWAN Configuration

LoRaWAN Parameter	**Value**
Modulation technique	LoRa (based on CSS)
SF	7
Coding rate	4/5
Bandwidth W	125 kHz
Transmission power	14 dBm
Center frequency f_k	$\{867.1, 867.3, 867.5, 867.7, 867.9, 868.1, 868.3, 868.5\}$ MHz

in Figure 4.9b. This Pycom node transmits an uplink packet for each specific channel f_k sequentially. While the gateway attempts to send an acknowledgment by default at the same frequency as the message transmitted. Accordingly, the node writes the information of the last received downlink packet (packet number, ESP, etc.) to the payload of the next uplink packet. A desktop computer runs a Python program is used as an AS which receives data from the LoRa Network Server (LNS), as well as LoRa metadata with all parameters of the LoRaWAN transmission (f_k, SF, Bandwidth W, RSSI, SNR, etc.). Consequently, the computer stores this data for processing as detailed in the following sections. For the research society, the data is provided on this online repository [106].

(a) The gateway connected to the antenna cable. (b) The end node at location 4. (c) The end node at location 5. (d) The end node at location 7.

Figure 4.9 – The gateway and views from the end node sites. For the end node, a packaged Pycom device fixed on a rod and connected to a battery is used.

4.4.4 Effective Signal Power

The LoRa system can operate at SNR below 0 dB as LoRa signals can be decoded with signal power below the noise floor [101]. Based on the independence between signal and noise, the measured power P_r at the input Radio Frequency (RF) chain is the sum of the signal power P_s, the potential interference power P_i and the noise power P_n. These are related on a linear scale as:

$$P_r = P_s + P_i + P_n \tag{4.10}$$

and the signal to interference plus noise ratio is:

$$\rho = \frac{P_s}{P_i + P_n}. \tag{4.11}$$

By assuming the power quantities are expressed in milliwatt, these can be redefined on a logarithmic scale as:

$$RSSI_{dBm} = 10\log_{10} P_r, \tag{4.12}$$

$$ESP_{dBm} = 10\log_{10} P_s, \tag{4.13}$$

and the SINR is:

$$SINR_{dB} = 10\log_{10} \rho. \tag{4.14}$$

Hence, the computation of the $RSSI$ follows as:

$$\begin{aligned} RSSI_{dBm} =& 10\log_{10}(P_s + \frac{P_s}{\rho}) \\ =& 10\log_{10}(P_s) + 10\log_{10}(1 + \frac{1}{\rho}) \\ =& ESP_{dBm} + 10\log_{10}(1+\rho) - 10\log_{10}(\rho) \\ =& ESP_{dBm} + 10\log_{10}(1 + 10^{\frac{SINR_{dB}}{10}}) - SINR_{dB}, \end{aligned} \tag{4.15}$$

then the ESP can be expressed as:

$$ESP_{dBm} = RSSI_{dBm} + SINR_{dB} - 10\log_{10}(1 + 10^{0.1 SINR_{dB}}). \tag{4.16}$$

By ignoring the potential interference power P_i, the RSSI lower limit can also be

defined with respect to the thermal noise power P_n as:

$$\begin{aligned} \underline{RSSI}_{dBm} =& 10\log_{10}(P_n\rho + P_n) \\ =& 10\log_{10}(P_n) + 10\log_{10}(\rho + 1) \\ =& 10\log_{10}(P_n) + 10\log_{10}(1 + 10^{0.1SINR_{dB}}) \end{aligned} \tag{4.17}$$

with

$$P_n = k_B T W \cdot 1000, \tag{4.18}$$

where the Boltzmann constant k_B is $1.381 \times 10^{-23}\ \mathrm{J\,K^{-1}}$ and the temperature T is set to 293.15 K. By using equation 4.16, the ESP lower limit can also be defined as:

$$\underline{ESP}_{dBm} = 10\log_{10}(P_n) + SINR_{dB}. \tag{4.19}$$

At this point, these can be estimated as:

$$\underline{RSSI}_{dBm} \approx -123 + 10\log_{10}(1 + 10^{0.1SINR_{dB}}) \tag{4.20}$$

and

$$\underline{ESP}_{dBm} \approx -123 + SINR_{dB}. \tag{4.21}$$

Taking the measured data in Rennes as an example, Figure 4.10 shows a comparison between RSSI and ESP against different SINR values. For positive SINR, ESP coincides with RSSI, but it differs for the negative SINR. The most striking observation is that the raw measurement of RSSI and ESP provided by the experimental setup are very precisely limited by expressions 4.20 and 4.21, while those lower limits are obtained by only considering the thermal noise assumption. Thus, the RSSI distribution saturates when the received power is approaching −120 dBm. In contrary to RSSI, ESP at negative SINR goes below this previous limited value of the RSSI. It is worth mentioning that when the ESP limitation gets to an extreme value, it refers to the maximum receiving sensitivity, i.e. −142 dBm as mentioned in the Tektelic gateway specifications sheet [104]. Since some packets are received with SINR below 0 dB in the practical experiments, ESP is more reliable to be used for CSI estimation and analysis, as detailed in the following sections. It could be utilized for IoT localization also, particularly ESP instead of RSSI

fingerprinting.

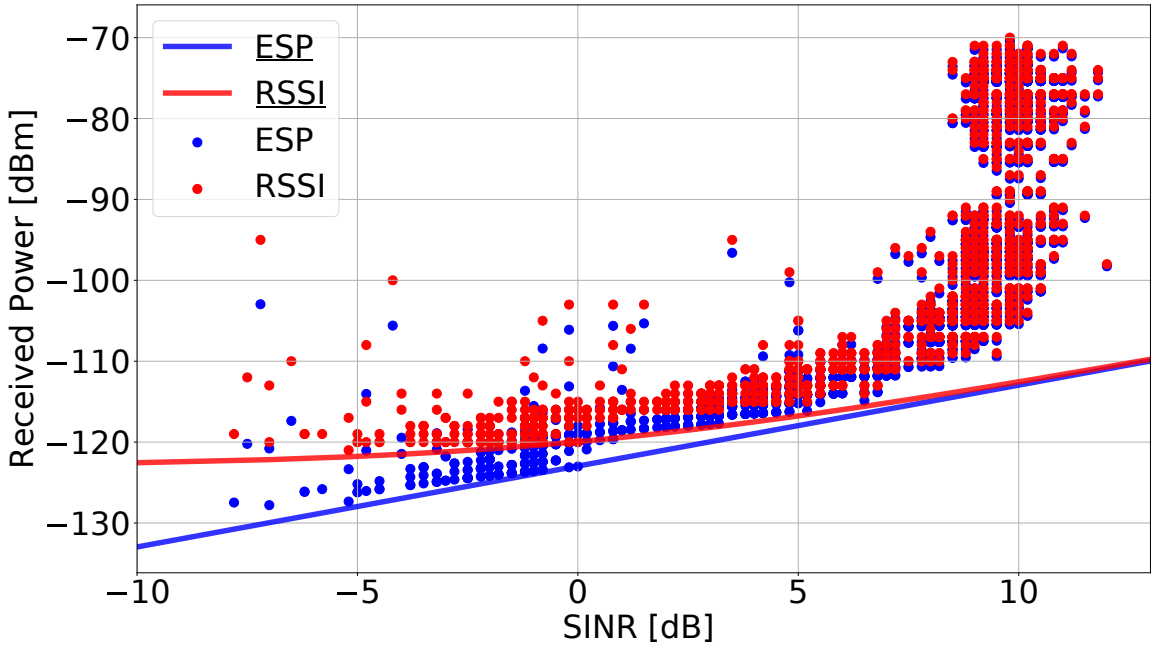


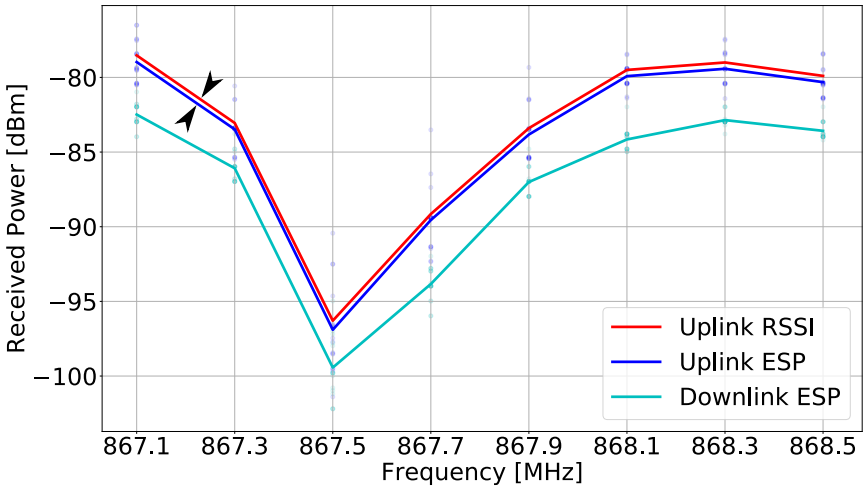
Figure 4.10 – Comparison of the measured RSSI and ESP against SINR.

4.4.5 CSI Estimation

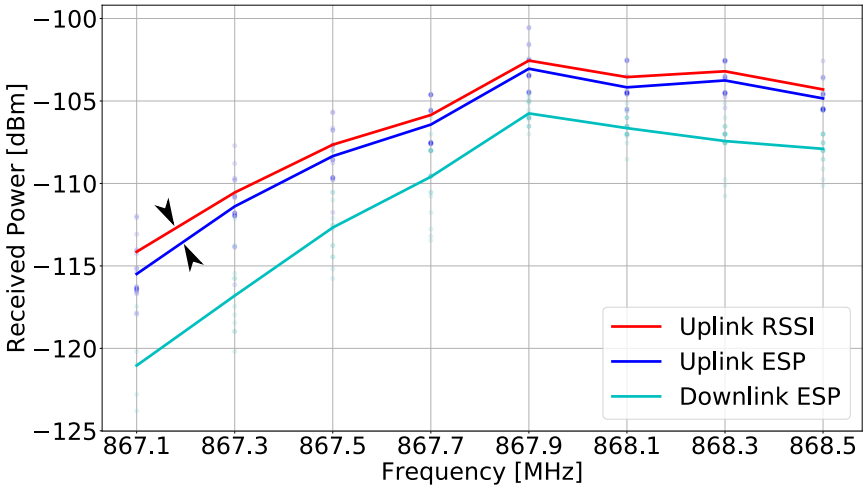
At the gateway side, the received data is extracted from the payload of each received packet plus the metadata. As an approach for obtaining the absolute CSI, uplink ESP, uplink and downlink ESP values are averaged at each frequency channel f_k. As the shown example in Figure 4.11a, the absolute CSI from location 2 is frequency selective with a deep fade of more than 15 dB depth, particularly at 867.5 MHz. Moreover, channel reciprocity is manifested between the uplink and downlink CSI with almost the same fading position. While in Figure 4.11b, the obtained CSI also indicates multipath propagation whose deep fades are up to 10 dB depth. Apart from a constant value whose compensation would require an accurate calibration of the transceiver RF chains, the same channel reciprocity across all the frequency bands is preserved.

On the other hand, one can observe that the gap between the uplink RSSI and ESP is larger in Figure 4.11b than in Figure 4.11a, as indicated by the black arrows. This

behavior is clarified by equation 4.16, thus, the gap between the RSSI and ESP increases when SNR decreases below 0 dB. Therefore, the RSSI values are ignored, whereas the ESP values are utilized for further investigations as detailed in the following sections.



(a) The absolute CSI from location 2.



(b) The absolute CSI from location 7.

Figure 4.11 – The absolute CSI shape from different locations.

4.4.6 Reciprocity of the CSI

In this section, the magnitude of the aforementioned CSI reciprocity is evaluated for all the locations. To quantitatively appreciate the similarity or dissimilarity of the frequency

dependency of quantities, for example, the uplink versus downlink ESP values, the PCC is used [107]. Let:

$$\mathbf{x} = [x_0, \ldots, x_{N-1}]^T \tag{4.22}$$

and

$$\mathbf{y} = [y_0, \ldots, y_{N-1}]^T \tag{4.23}$$

be two frequency dependent vectors of the same size N, where N in this experiment corresponds to the number of the frequency channels f_k. Consequently, the centered versions of those two vectors are:

$$\mathbf{x_c} = \mathbf{x} - \frac{\mathbf{x}^T.\mathbf{1}_N}{N} = \mathbf{x} - \bar{\mathbf{x}} \tag{4.24}$$

and

$$\mathbf{y_c} = \mathbf{y} - \frac{\mathbf{y}^T.\mathbf{1}_N}{N} = \mathbf{y} - \bar{\mathbf{y}}. \tag{4.25}$$

At this point, the PCC between those two vectors is classically expressed as a simple dot product between their normalized versions as:

$$\rho_{xy} = \hat{\mathbf{x}}_\mathbf{c}^T.\hat{\mathbf{y}}_\mathbf{c} = \frac{{\mathbf{x_c}}^T.\mathbf{y_c}}{\sqrt{{\mathbf{x_c}}^T\mathbf{x_c}}\sqrt{{\mathbf{y_c}}^T\mathbf{y_c}}}. \tag{4.26}$$

This correlation coefficient is a measure of the linear correlation between those two vectors which captures only their fluctuations around their centers. Hence, this coefficient is essentially a normalized value, such that the result always has a value between -1 and 1.

By utilizing the PCC for quantifying the reciprocity of the uplink versus downlink CSI, Figure 4.12 shows the magnitudes of the PCC across all the locations. The result shows that most of the PCC values are greater than 0.9 which represents a nearly identical shape of CSI in the uplink and downlink. For future IoT applications, this reciprocity feature could be exploited for the physical layer security between the end node and the gateway [100]. However, location 3 and 10 have lower PCC values due to the lack of the averaged ESP values at some frequency bands, hence, this often happens as a reason of the packet transmission losses.

4.4.7 Spatial Correlation of the CSI

For investigating the spatial correlation between the CSI shapes from different positions, the CSI spatial evolution at each location is also analyzed using the aforementioned

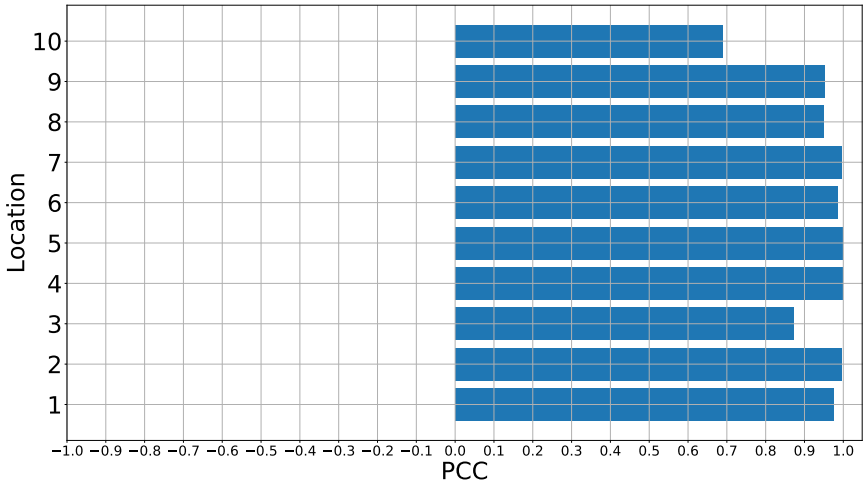
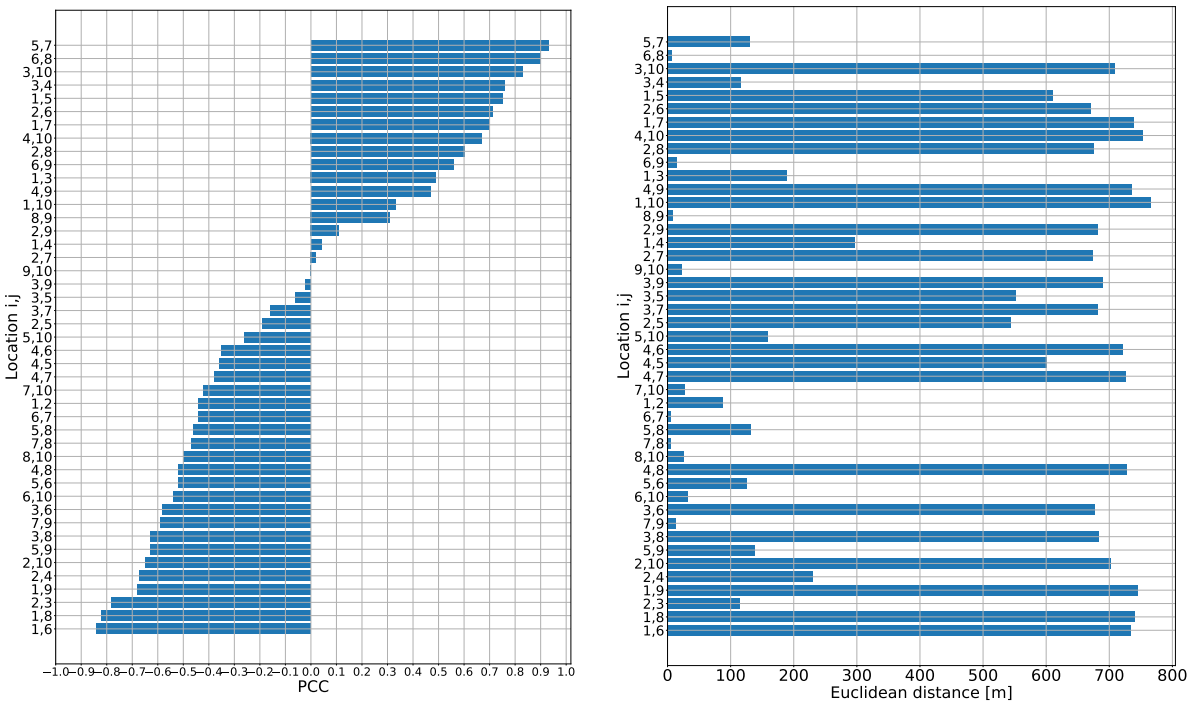


Figure 4.12 – The PCC magnitude between the uplink and downlink CSI at each location.

method based on the PCC between each location combination i and j. As depicted in Figure 4.13a, most of the PCC magnitudes are generally unrelated to the Euclidean distance at any two locations in Figure 4.13b. In contrary to the nearly similar uplink and downlink CSI which is demonstrated by the high reciprocity for a single stationary location, hence, the CSI shape is varying from one place to another even for short separations. This behavior is reasonable as the minimum Euclidean distance between any two locations is $\approx 4.9\,\mathrm{m}$ in this experiment. This suggests the feasibility of utilizing the CSI for localization provided this CSI remains stationary enough over time. Thus, the accuracy of positioning could be improved by utilizing the rich channel information in each subcarrier f_k as well as the frequency hopping in the LoRa systems. Hence, the individual CSI from each different location has to be appropriately different from one another to achieve significant localization gain.

4.4.8 Conclusion

This section investigates both the spatial evolution of the CSI from different locations, as well as, the reciprocity between the uplink and downlink CSI in LoRa is analyzed. Accordingly, a measurement campaign is carried out in the city of Rennes to estimate the CSI by averaging ESP values at each frequency band. Hence, ESP is more reliable than RSSI as it overcomes the RSSI limitation, especially at low SNR ($< 0\,\mathrm{dB}$). Thus, PCC is utilized to measure the linear correlation between two CSI, as a result, the uplink and



(a) PCC values.

(b) Euclidean distance values.

Figure 4.13 – PCC and Euclidean distance values across all the combinations.

downlink CSI are almost analogous at most of the locations. Furthermore, the CSI shapes from different locations are highly uncorrelated from each other.

For future work, this section recommends using this reciprocity feature for the physical layer security between the end node and the gateway. Moreover, the unique CSI shape at each location could achieve significant localization gain by utilizing the frequency hopping in the LoRa systems to get the rich channel information in each subcarrier.

4.5 Summary and Conclusion

This chapter investigates the feasibility of localizing using LoRa CSI, by performing outdoor measurement campaigns in the area of Beaulieu Campus in Rennes. Thus, the spatial and temporal stability of CSI are asserted to be relatively stronger than the RSSI values. At each different location, the CSI reciprocity is quantified using the PCC which shows a very high linear correlation between the uplink and downlink CSI. On the other hand, most of the CSI shapes from different locations are highly uncorrelated to each other. Hence, it can be anticipated that this could achieve significant localization gain by utilizing the frequency hopping in the LoRa systems by getting access to a wider band. In the given results, the feasibility of using the ESP is proven by its enlarged range when the SNR is very low, unlike the RSSI which has a limitation. Without loss of generality, LoRa CSI can be obtained using the following proposed approaches:

1. **Estimating the CSI from the IQ samples** of the received LoRa preamble chirps at the gateway. For each center frequency, a narrowband CSI is estimated until the whole CSI shape is acquired by the frequency hopping, as shown in Figure 4.14.
2. **Estimating the CSI by averaging the ESP values** for each center frequency until the whole CSI shape is acquired by the frequency hopping. This approach is considered the most feasible one as it doesn't require any additional hardware on the gateway side. To obtain the ESP value, the value of RSSI and SNR is the only requirement that is mostly available in any traditional gateway.

For future work, this chapter highly recommends using ESP for the potential IoT applications, especially localization using ESP fingerprinting instead of using RSSI. Thus, the different channel gains can be used as additional information to increase the localization accuracy.

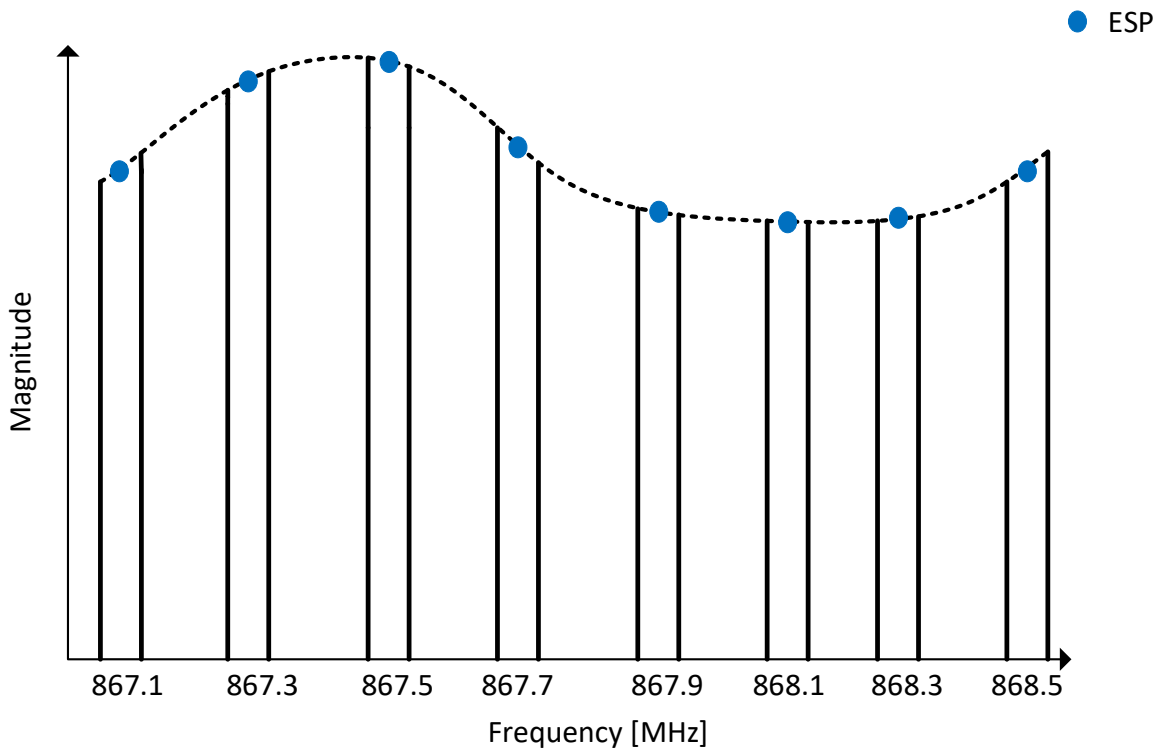


Figure 4.14 – Estimating the wide-band CSI shape by either integrating the narrow-band CSI shapes or the ESP values of the frequency bands.

Part II

Spectrum Allocation for IoT Networks

Chapter 5

Characterizing and Modeling of PDR in an Actual LoRaWAN Network

5.1 Introduction

With the growth of the IoT, it is critical to understand the packet transmission performance in LPWAN, especially LoRaWAN. Hence, the transmission quality of data packets has a significant impact on the performance of the entire LoRaWAN [15]. Thus, the reliability of communications in the network is impaired by the channel quality of the link between the end node and the gateway. Consequently, packet loss may occur due to the channel attenuation or interference. This transmission failure affects various IoT applications and even causes serious outcomes. Furthermore, the data integrity and accuracy in IoT data analytics may be reduced by a large amount of packet loss. Therefore, LoRaWAN supports both unconfirmed and confirmed messaging. Confirmed messaging is most probably used for important sensor data. In contrary to the unconfirmed message, the end node requires the message to be acknowledged as received by the network server when sending a confirmed message. Accordingly, the end node will retransmit the data packet when it does not receive the acknowledgment. However, this retransmission requires an extra energy consumption that impacts the battery life of end node devices plus occupying an additional spectrum, raising the interfering level and occupying time that could be used for uplinks.

5.2 Related Works

Recently, many papers are addressing the different origins of LoRa packet loss. For example, in [101] the authors conduct experiments to evaluate a characterization of LoRaWAN frame collision conditions. Moreover, the PDR and coverage of LoRa technology are evaluated for outdoor cases in [96] and [14]. Based on empirical results in [98], path loss models are developed for LoRaWAN communications and compared with widely used empirical models. In [108], authors study the impact of environmental factors on the performance of LoRa, and show that higher temperatures decrease the RSSI and may drastically affect PDR. While in [13], the first attempt to investigate the PDR in a real city-scale LoRaWAN network is presented. However, these previous works do not evaluate the ESP, i.e. introduced in Section 4.4.4, as one of the main factors affecting the PDR rather than evaluating RSSI only. Besides, they do not focus on the frequency dependency of PDR. Over and above, these previous works don't introduce the modeling of PDR corresponding to the received power. Hence, the received power is considered as one of the main factors affecting the PDR.

5.3 Contribution

In this chapter, an in-depth investigation of the frequency dependency of the PDR is done by utilizing the measured data of a measurement campaign, i.e. introduced in Section 4.3.3. From each different location, the ESP and SINR values are obtained as well as the influence of these parameters on the PDR is evaluated at each frequency band independently. Hence, the whole ESP values in the experiment are extensively evaluated, and the other possible causes of packet loss are thoroughly explained. With the given results, the feasibility of using the ESP is proven by its enlarged range when the SINR is very low, unlike the RSSI which has a limitation. Accordingly, modeling the PDR using an ESP parameterized beta distribution function is proposed. Feasibility of the proposed model is assured by simulating PDR against ESP, hence, the simulated PDR values follow the distribution of the measured ones well. This PDR model gives important guidelines for future LoRaWAN network regulation and optimization. Moreover, this investigation manifests and gives important guidelines for using ESP in the future IoT applications.

5.4 Organization

The remainder of this chapter is organized as follows. Section 5.5 presents the main factors of the packet losses. Section 5.6 provides an evaluation of the PDR's frequency dependency. The effect of the channel quality on PDR are then presented and commented in Section 5.7. Hence, Section 5.8 provides the modeling of PDR from the measured data. Finally, the work is concluded in Section 5.9.

5.5 Key Factors of Packet Loss

Packet losses are the main drawback for IoT [21]. Hence, packet losses cause many retransmissions at the cost of a lower battery lifetime of the end nodes and may lead to an increase of the RF contention level. Even worse, a total failure of the IoT service could happen, either because end nodes can not succeed in sending any data to the gateway or because all their energy is consumed much faster than expected due to the multiple repetitions of the transmission. Without loss of generality, LoRaWAN is used as an example in this thesis but any other IoT protocol could be utilized. Thus, the major sources of the packet loss are described in the following subsections, as shown in Figure 5.1.

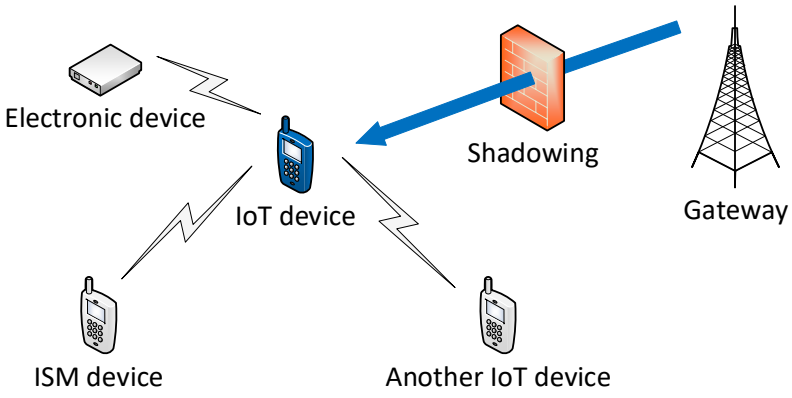


Figure 5.1 – An envisioned IoT network with the potential factors of the packet losses.

5.5.1 Packet collision

Channel contention occurs when multiple devices attempt to send data over the same channel simultaneously [10, 109]. Whereas the end nodes are uncoordinated, the packet transmission is initiated by the end node, not the network. Hence, the traditional LoRa device does not check if the channel is preempted by other devices before transmitting a packet, so it may cause packet collision. Additionally, the IoT networks are superposed with no coordination between them. So, these collisions could also occur no matter whether the interfered signal is from other end nodes of the same network or the surrounding IoT networks, using the same IoT standard or not. Moreover, interference could occur from other radio signals present in the ISM bands which are not IoT signals, so they can be considered "jammers" by definition. Last but not least, electromagnetic radiations could disturb the end nodes due to the proximity of other electronic devices suffering from leakage radiations, for instance, this could happen in an industrial environment.

5.5.2 Channel factors

The end node can be located far away from the gateway, especially in LPWAN networks, which causes shadowing and signal attenuation across the transmission range [98]. Packet transmission success is closely related to the channel state, indeed, a channel facing bad propagation conditions has the same effect as caused by collisions. These propagation conditions are directly determined by the channel parameters which are SNR, RSSI and ESP. SNR is the ratio of signal power to the measured noise power. Consequently, the higher SNR is, the smaller the noise mixed in the signal, and the easier it is to separate the effective signal. RSSI is a relative value of signal power measured by the end node or the gateway at its receiving end. However, ESP is a more reliable parameter as it overcomes the RSSI limitation, especially at low SNR ($< 0\,\text{dB}$) as depicted in the previous Section 4.4.4.

Without loss of generality, ESP is computed from $SNR \mid SINR$ as depicated in Equation 4.16. In general, ESP value decays with the shadowing effect and increase of distance, in particular for the long transmission range, as in most IoT applications. Due to the low transmission power, i.e. maximum transmit power of the LoRa signal is 14 dBm, and the large propagation attenuation, this value is usually negative in dBm. Furthermore, these channel parameter values depend on the specific environmental conditions around

each end node, hence, each CSI is different from one location to another.

5.6 Frequency Dependency of PDR

The radio channel between a gateway and an IoT node in a LoRaWAN is very specific at each location. Its quality has an impact on the capability to establish a successful transmission, as it is confirmed in the previous sections. Moreover, the LoRa channel with such a small bandwidth of 125 kHz is almost flat. However, this section shows that there are significant variations of the channel gain over different LoRa center frequencies f_k, with only a 200 kHz spacing in the 868 MHz ISM band. Consequently, the impact of the channel on the successful packet rate is investigated for each center frequency independently. In an offline mode, the received data is extracted from the payload of each LoRa packet at the gateway side. Thus, ESP value is extracted and analyzed against the PDR value at each frequency channel f_k. On the other hand, PDR is calculated at each different channel f_k independently as:

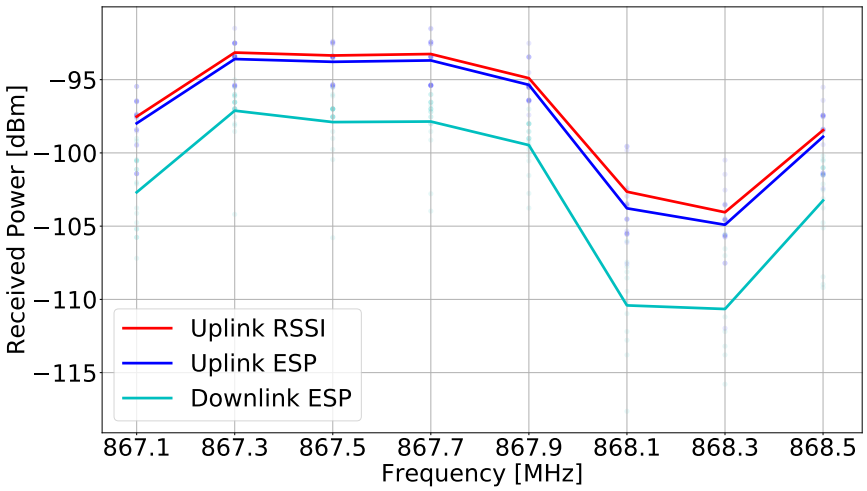
$$PDR = \frac{\text{Number of received packets}}{\text{Number of transmitted packets}}, \tag{5.1}$$

where the number of transmitted packets to obtain the PDR value at each frequency band f_k is 20.

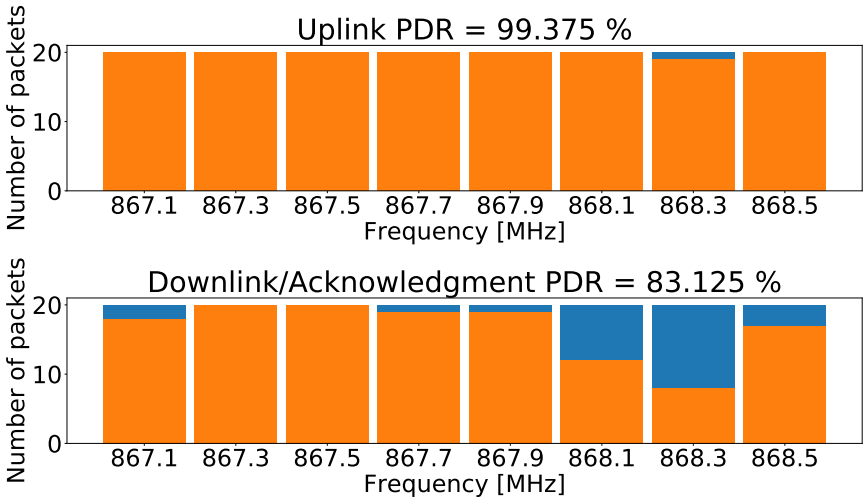
As an example shown in Figure 5.2a, the absolute CSI from location 4 is frequency selective with a deep fade of more than 10 dB depth. On the other hand, channel reciprocity is clearly manifested between the uplink and downlink CSI apart from a constant value whose compensation would require an accurate calibration of the transceiver RF chains. While the PDR values preserve almost the same magnitude ranks across all the frequency bands, as shown in Figure 5.2b. Here, downlink PDR values are preferred because they are more correlated with the absolute CSI shape in these near transmission conditions.

5.6.1 CSI shape variability over the ISM-Band

Figure 5.3a shows exemplary signal level spectra observed in all the scenarios of the experiment, averaged over the whole trial. A progressive reduction is observed in the received power across location 1 to 10. This reduction is reasonable as the distance between the gateway and the node gradually increases from $\approx$ 5 m at location 1 to $\approx$ 760 m at location 10. The first category of CSI is represented in location 1, 3, 8, 9, and 10 which



(a) ESP values.



(b) PDR values.

Figure 5.2 – Result from location 4 at each different frequency band.

shows moderate fades with a low indication of multipath propagation. Thus, the variation of signal level is less than 6 dB over the 1.5 MHz frequency range. The opposite is the signal level spectrum of location 2, 4, 5, 6, and 7 which stand for another category of channels showing strong frequency selectivity. This category indicates multipath propagation with a path delay so it has many deep fades with up to 15 dB depth, in such a small bandwidth of 1.5 MHz only.

5.6.2 Impact of the channel quality on PDR

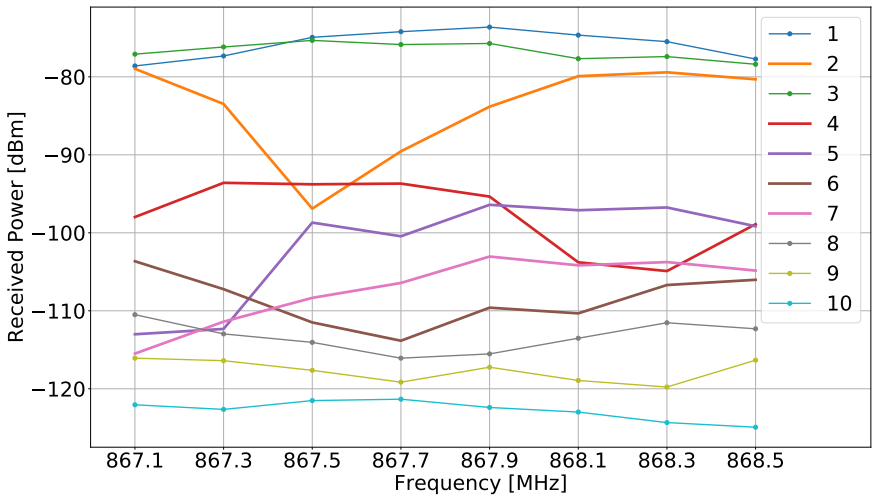
The PDR values on each center frequency for all the locations are shown in Figure 5.3b. Location 1 and 3 show total successful transmissions across the whole frequency bands which is reasonable as they already have almost flat fading channels, and high received power as shown in Figure 5.3a. While PDR values in the other locations preserve the same pattern of their CSI shape. For example, location 5 has PDR values that fit well with respective properties of the corresponding frequency bands, particularly at 867.1 MHz and 867.3 MHz with low PDR values like their ESP values. This confirms the strong frequency dependency of the PDR in the LoRa system.

5.7 Impact of channel parameters on PDR

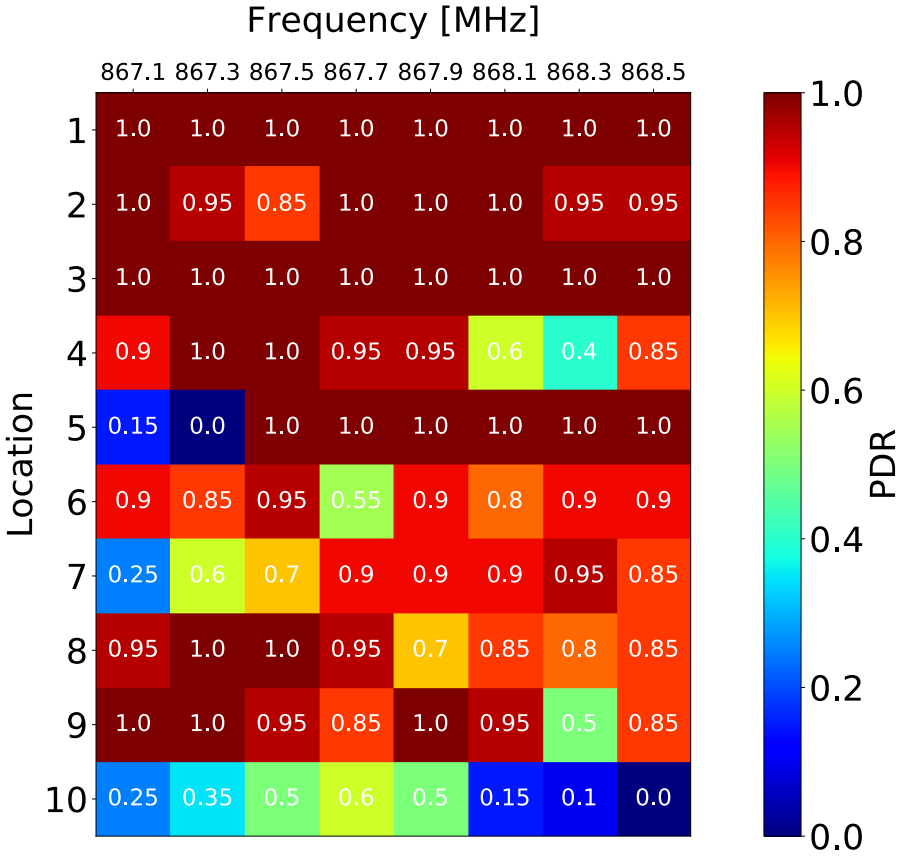
To check the plausibility of utilizing the ESP, the main causes of the packet loss events are discussed in this section. Two main aspects can affect the PDR, which are measured received power and SINR, as detailed in the following subsections.

5.7.1 Impact of received power

Figure 5.4a explores how downlink PDR changes along with uplink RSSI and ESP. Hence, one can observe that the PDR magnitudes are uniformly quantized with a resolution of $\frac{1}{20}$. Nevertheless, there are some irregular gaps between the PDR values which could be compensated by taking more measurement data. On the other hand, the result shows that high PDR values reduce significantly with the decrease of RSSI, which is similar when the ESP decreases. However, RSSI distribution is limited when approaching the aforementioned noise floor limitation, i.e. $RSSI \approx -120\,\text{dBm}$ at $SINR = 0\,\text{dB}$. One can observe the rather large gap between the ESP values and the RSSI values, especially when



(a) ESP values over different frequency bands at each location.



(b) PDR values over different frequency bands at each location.

Figure 5.3 – Result from the ten different locations.

the received power is below −120 dBm. Therefore, ESP has a simpler and more natural relationship with PDR than RSSI.

5.7.2 Impact of SINR

The result shows a general rise in PDR when the color-coded ESP or SINR increases, as shown in Figure 5.4b. This direct proportionality is established in the previous works [13]. On the other hand, it can be noticed that this long-term observation exhibits two kinds of unexpected behavior. These should be related in two manners in how the interference can affect the relationship between received power and SINR. The first unexpected situation is to have at the same time a low SINR and a rather high received power, as depicted in Figure 4.10. Those situations correspond to a case where the SINR is probably measuring an additional interfering power coming from a colliding packet. The second unexpected situation is having a high SINR and nevertheless, a low packet success rate as depicted in Figure 5.4b, which is most probably due to a typically accomplished interference in the downlink. Those observations suggest that the joint observation of SINR and RSSI or ESP can be exploited for detecting interfered conditions.

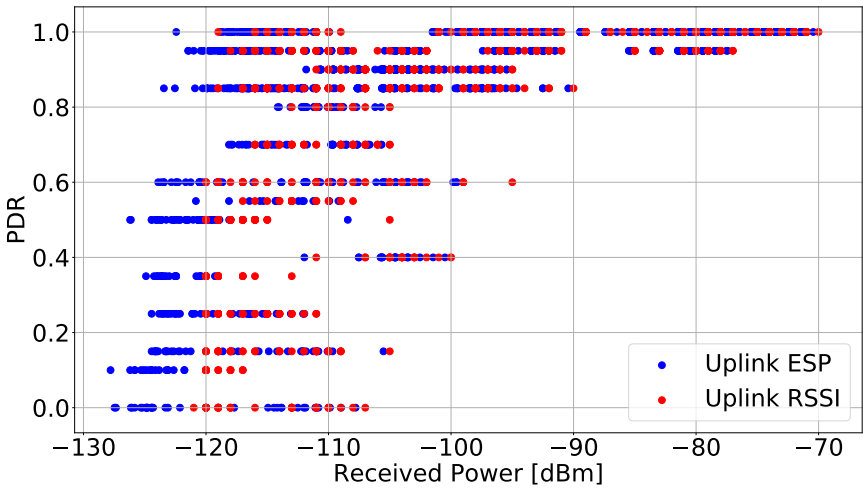
5.8 Modeling of PDR as a function of ESP

5.8.1 Contribution

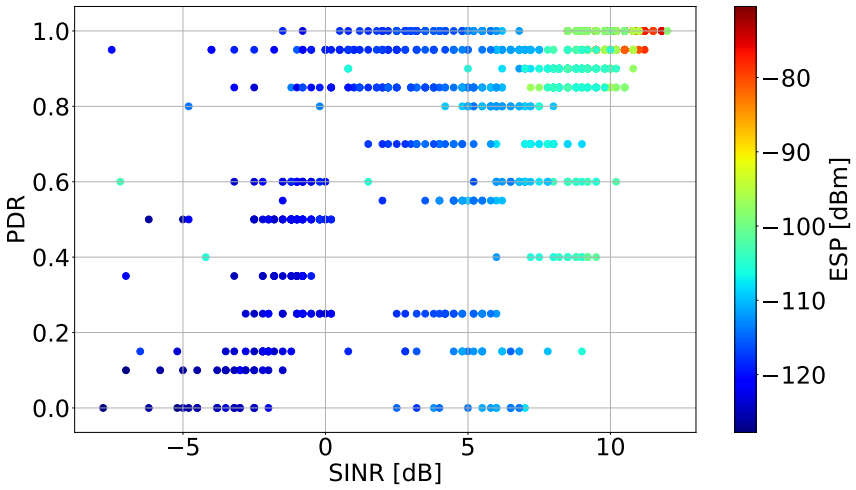
In this section, an in-depth investigation of the PDR using the traces collected from the real network deployment of the aforementioned measurement campaign. Hence, the whole ESP values in the experiment are extensively evaluated, and then a stochastic model is proposed for modeling PDR as a function of the ESP rather than using RSSI to overcome the RSSI's limitation at low SNR (< 0 dB).

5.8.2 Organization

The remainder of this section is organized as follows. Section 5.8.3 provides sufficient detail about the PDR modeling based on the measurement's result. Hence, Section 5.8.4 provides an evaluation of the simulated PDR based on the proposed model. Finally, the work is concluded in Section 5.8.5.



(a) Changes with the measured received power.



(b) Changes with the measured SINR.

Figure 5.4 – PDR for various received power and SINR values.

5.8.3 PDR Modeling using Beta distribution

For system-level simulations or for dimensioning actual IoT deployment performances, it is very useful to benefit from a PDR model that encompasses all the possible situations and their very large observed variability. However, the PDR values show extreme variations which make the exploitation of their empirical mean difficult, as shown in Figure 5.5a. Therefore, the first step of the proposed approach is to get access to a reasonable estimation of the expected PDR average by logistic regression. This regression takes into account all the datasets at once by fitting a smooth and reasonable curve to the measured values. Consequently, the mean of the PDR distribution can have a deterministic model using a logistic function or logistic curve which is a common sigmoid curve with an equation as:

$$\mathbb{E}[PDR] = \frac{1}{1 + e^{-k(ESP - ESP_0)}}, \tag{5.2}$$

where k is the logistic growth rate or steepness of the curve which is estimated to be 0.238. While the ESP value of the sigmoid's midpoint ESP_0 is $-119.794\,\text{dBm}$. These parameters are estimated while setting SF equal to 7 during the experiment, hence, these values can be different if another configuration is set. On the other hand, the expected PDR values lie around an interval that can be sized with a given level of confidence 2σ. This σ value is obtained based on the number of packets used in this experiment to produce the empirical estimation of each PDR value. This confidence region is the shaded blue region in Figure 5.5a, accordingly, the PDR values are estimated roughly around this region.

In a second step, in order to simulate the observed huge variability in the measured PDR, an ESP parameterized Beta distribution of PDR is proposed [110]. Consequently, the Probability Density Function (PDF) of the PDR distribution can have a stochastic model using a Beta distribution function as:

$$\begin{aligned} f(PDR; \alpha, \beta) =& \frac{1}{B(\alpha, \beta)} PDR^{\alpha-1}(1 - PDR)^{\beta-1} \\ =& \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} PDR^{\alpha-1}(1 - PDR)^{\beta-1} \end{aligned} \tag{5.3}$$

with

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx, \tag{5.4}$$

where $\Gamma(z)$ is the gamma function. While the beta function $B(\alpha, \beta)$ is a normalization

constant to ensure that the total probability is 1. In the above equation, PDR is a realization, i.e. an estimated PDR value that actually occurred, of a random process. Moreover, this Beta distribution function $f(PDR; \alpha, \beta)$ has a mean $\mathbb{E}[PDR]$ which is computed as:

$$\mathbb{E}[PDR] = \frac{\alpha}{\alpha + \beta}. \tag{5.5}$$

By exploiting the noticeable formal similarity between the two expression of the expectation in equation 5.2 and 5.5, α and β are estimated as $\alpha = 1$ and $\beta = e^{-k(ESP - ESP_0)}$. At this point, these can also be redefined as:

$$\mathbb{E}[PDR] = \frac{1}{1 + \beta} \tag{5.6}$$

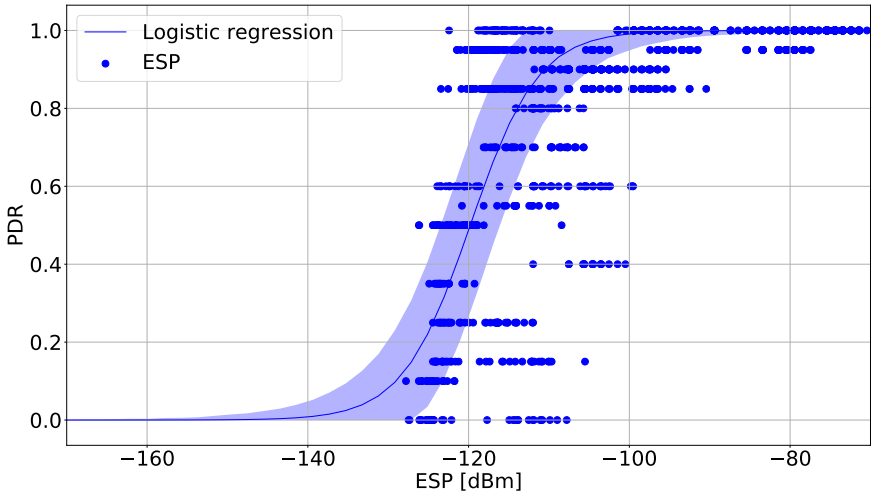
and the variance $\mathrm{Var}[PDR]$ is:

$$\mathrm{Var}[PDR] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{\beta}{(1 + \beta)^2(\beta + 2)}. \tag{5.7}$$

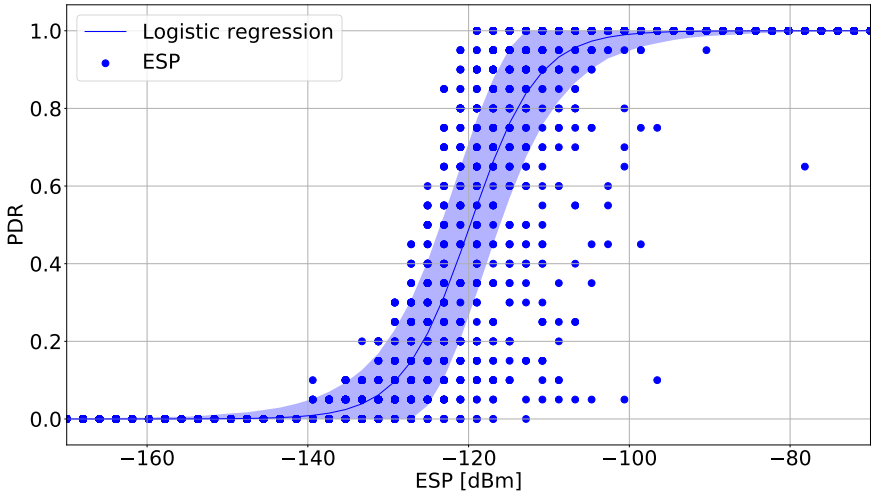
5.8.4 Simulation Results

In this section, the performance of the proposed PDR model is evaluated by generating PDR distributions at each ESP value. Figure 5.5b shows the simulated PDR distribution which follows the distribution of the measured PDR in Figure 5.5a. Consequently, the expectation of the simulated PDR values lies within the 2σ confidence interval around the sigmoid curve, whereas the extreme variations of the PDR are well simulated as promised by this proposed model. For obtaining the same resolution of the measured PDR, the simulated PDR values are uniformly quantized with the same resolution of the measured PDR.

On the other hand, the PDR distribution is different at each ESP value as manifested by their corresponding Beta distribution functions in Figure 5.6. For example, the Beta distribution function at $-130\,\mathrm{dBm}$ has a distribution concentrated around low PDR values, in contradiction with the distribution at $-110\,\mathrm{dBm}$ which has high PDR values. While the distribution at $-119.794\,\mathrm{dBm}$, i.e. sigmoid's midpoint ESP_0 as depicted in equation 5.2, is considered the most ambiguous one which is an almost uniform distribution. In this uniform case, the packet delivery rate has an equivalent probability of success and failure.



(a) Measured PDR versus different ESP values.



(b) Simulated PDR versus different ESP values.

Figure 5.5 – PDR against different ESP values.

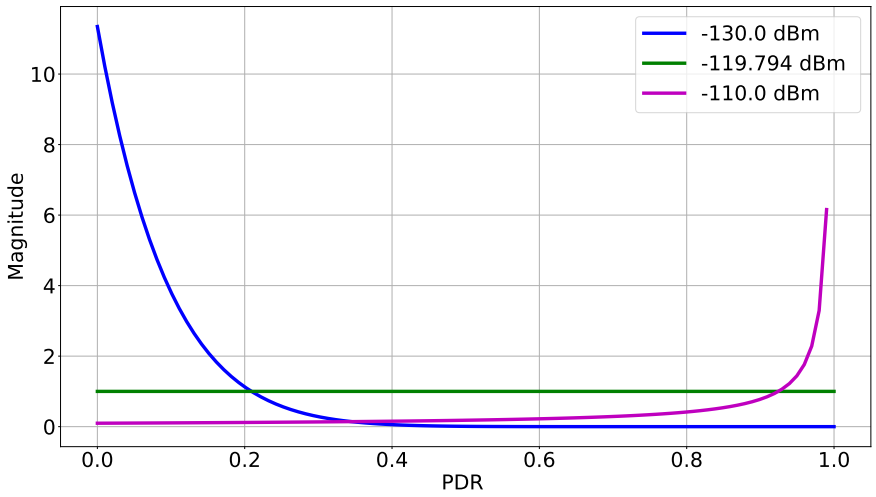


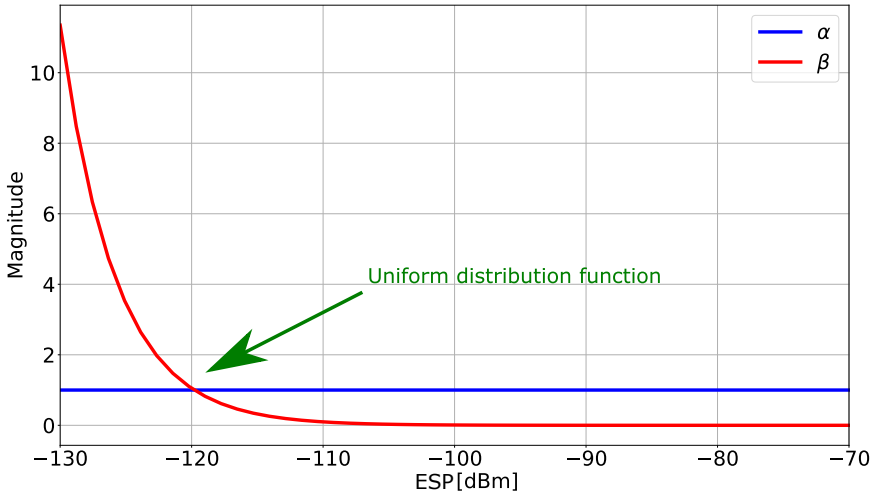
Figure 5.6 – Beta distribution function $f(PDR; \alpha, \beta)$ at different ESP values.

Moving to the fine parameters of the Beta distribution function $f(PDR; \alpha, \beta)$. Figure 5.7a shows the constant α and the variable β across different ESP values. One can observe that as β is changed, the shape of the distribution changes. As α becomes larger than β (more successful packets), the bulk of the probability distribution will shift towards the high PDR values as shown in Figure 5.5b, whereas an increase in β moves the distribution towards the low PDR values (more failures). While α and β are equal to 1 at $-119.794\,\text{dBm}$ which is depicted by a green arrow and as a uniform distribution function in Figure 5.6.

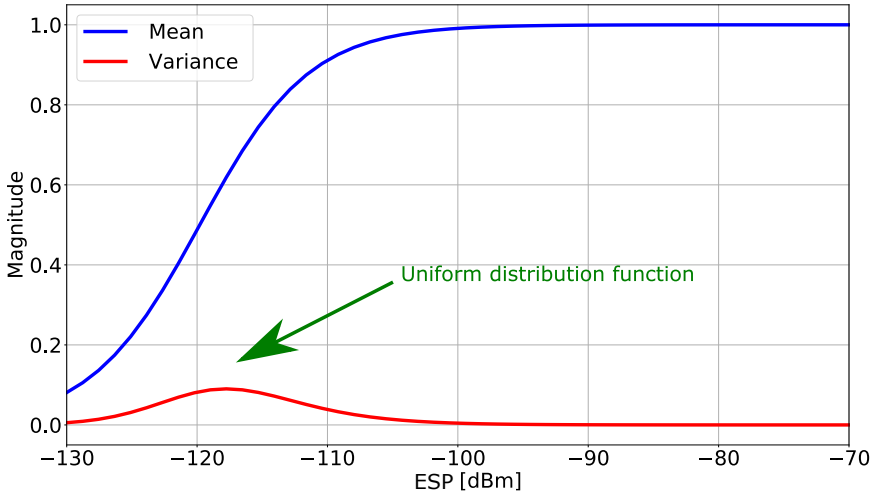
Focusing on the introduced mean $\mathbb{E}[PDR]$ and variance $\text{Var}[PDR]$ of the Beta distribution function, Figure 5.7b showcases their values with a wide range of ESP values. Generally, as ESP increases, the mean increases until it starts to saturate when ESP is above $-100\,\text{dBm}$. While the variance is very small at most of the ESP values, however, it has a peak when the ESP is around $-119.794\,\text{dBm}$ as shown by the green arrow. This peaky region is indicated in Figure 5.5b when PDR distribution has a large variance around $-119.794\,\text{dBm}$.

5.8.5 Conclusion

ESP is considered to be more reliable than RSSI as it overcomes the RSSI limitation, especially at low SNR ($< 0\,\text{dB}$). Accordingly, this section presents the modeling of the



(a) α and β evolution against ESP.



(b) Mean $\mathbb{E}[PDR]$ and variance $\mathrm{Var}[PDR]$ evolution against ESP.

Figure 5.7 – Parameters of the Beta distribution function $f(PDR; \alpha, \beta)$.

PDR as a function of ESP. From the data of an accomplished measurement campaign, the measured PDR is investigated against ESP and is modeled using a Beta distribution function. Besides, the PDR distribution is simulated using the proposed model, as a result, the simulated PDR follows the measured one well. Furthermore, the fine parameters of the Beta distribution function are tuned at each ESP value for analyzing their effect on the PDR distribution.

For future work, this section recommends using this Beta distribution function for modeling PDR as a function of ESP. Thus, the potential studies in IoT can simulate the PDR against ESP to have an estimation for dimensioning the LoRaWAN network.

5.9 Summary and Conclusion

In this chapter, the frequency dependency of PDR in LoRa is studied. Thus, the channel parameters (RSSI, ESP, SINR) from different node locations are estimated during a measurement campaign in the city of Rennes. Hence, PDR is investigated against ESP at each different frequency. At each different location, the CSI shape is proven to be correlated with the PDR values. On the other hand, the presented data indicate that ESP has a more natural relationship with PDR than RSSI. Based on that, PDR is modeled as a function of ESP, hence, this proposed model could be utilized for future LoRaWAN network regulation and optimization.

As detailed in the next chapter, this frequency dependency of ESP should be exploited to improve PDR.

Chapter 6

Decentralized Adaptive Spectrum Learning in Wireless IoT Networks based on Channel Quality Information

6.1 Introduction

With the growth of the IoT usages in various applications, packet losses are increasing due to channel impairments on the link between the end node and the gateway, as detailed in the previous chapter. Therefore, it is essential to reduce the packet losses, especially in LPWAN. Decentralized artificial intelligence techniques have been proposed against collisions, but the approach here is extended to cope with the propagation disturbances. In this chapter, a QoC-A learning technique based on bandit algorithms is proposed in order to choose the transmission channel, so that it mitigates the propagation impairments between the radio channels using the channel quality parameter, i.e. ESP. Moreover, a DQoC-A algorithm is proposed to adapt rapidly to any abrupt change in the channels' conditions. A real experiment campaign is performed in the city of Rennes to demonstrate the low complexity and the feasibility of these proposed algorithms while implementing them on the IoT device side (decentralized approach). In the given results, QoC-A outperforms the classical UCB policy with a more accelerated learning process. On the other hand, the feasibility of using the DQoC-A in non-stationary scenarios is illustrated by its rapid convergence when abrupt changes in the channels' conditions occur. At the end of the process, these proposed learning techniques are assured to give fewer packet losses than the state-of-the-art ones with a random frequency allocation. These proposed reinforcement learning algorithms are recommended to be implemented in the end node side, moreover, important remarks are given for adjusting the algorithms' configurations based

on the properties of each of the potential IoT application.

6.2 Related Works and Motivation

Recently, many works are proposing different solutions for reducing the packet loss rate. For example, in [111] the authors propose DyLoRa, a dynamic LoRa transmission control system to improve energy efficiency, as an alternative to the state-of-the-art LoRaWAN ADR. Thus, the idea of DyLoRa is to adjust the transmission parameters, i.e. transmission power and SF, among different environments. Moreover, a Collision Avoidance Resource Allocation (CARA) algorithm is proposed in [112] with the objective of increasing the system capacity by mitigating packet losses from the collisions. However, these previous works are either centralized approaches (on the network side), hence, they do not consider the different interference conditions in the proximity of each end node which is located in a specific area. Or, they are imposing changes on the IoT protocol with extra packet transmissions and time synchronization between the end nodes. While in [113] and [109], the first implementation of a decentralized spectrum learning for IoT wireless networks is proposed in order to mitigate radio collisions without changing anything to LoRaWAN protocol. Moreover, they can be used in addition to conventional ADR techniques. They propose to use a learning algorithm on LoRa devices that can cope with spectrum scarcity which could occur in unlicensed bands. Nevertheless, this learning algorithm does not exploit the channel quality at each frequency band on the end node side to accelerate the learning process. This decentralized approach could be extended, which was just to avoid collisions to the management of channel quality too. Hence, the channel quality is estimated from the channel parameters, i.e. SNR, RSSI and ESP, which are considered as one of the main factors affecting the packet loss rate [11, 12]. Moreover, these channel parameters could be obtained on the end node side for implementing a decentralized approach, thanks to the reciprocity between the uplink and downlink CSI.

6.3 Contribution

In this chapter, a machine learning algorithm is proposed for spectrum allocation which could be embedded at the IoT end node side with low hardware extra cost in terms of processing power, memory footprint, etc. This algorithm reduces the packet losses by

learning a proper frequency allocation schema based on ESP as the channel quality indication to avert using a channel with poor propagation conditions and accelerating the learning process. Moreover, an adaptive algorithm is also proposed for non-stationary scenarios when the end node moves from one place to another or alternatively the CSI shape changes. Another non-stationary example could be when the end node exhibits a changing environment model, for example, if it is placed in a road where there are moving cars and objects. These moving obstacles may introduce a temporal evolution of the CSI shape between the morning and night period [114]. In a real network deployment in the university campus in Rennes, these algorithms are implemented inside an end node device and executed in different scenarios, i.e. stationary and non-stationary. Consequently, the whole results from the experiment are extensively evaluated among the proposed reinforcement learning algorithms against the state-of-the-art ones with random frequency allocation.

6.4 Organization

The remainder of this chapter is organized as follows. Section 6.5 provides sufficient detail about the system model of the proposed algorithm. The proposed algorithms are then illustrated in Section 6.6. While Section 6.7 shows the experimental architecture and network configuration used for the measurement campaign. In Section 6.8, the experimental results of the proposed algorithms are tested through the different scenarios. Additional remarks are given for adjusting the algorithms' configurations in Section 6.9. Finally, the work is concluded in Section 6.10.

6.5 System Model

The proposed learning approach can help IoT devices to reduce packet losses due to both weak channel propagation conditions and also collisions. That for, the proposed approach is inspired from [113] and [109] where a solution to radio collisions is proposed, while imposing no change on the IoT protocols, as, for instance, LoRaWAN. Employing the acknowledged messaging mode is the only required condition to utilize the proposed solution. Based on the investigation in [11], the underlying hypothesis is that the channels' PDR is not equally balanced across the K different frequency bands an IoT device can use. In other words, some channels are less attenuated than others. These channel con-

ditions are possibly be predicted online in time and space in a decentralized manner (on the end node side). As the end nodes may be quite far away from the gateways and suffer from different jamming and channel conditions, it is much more efficient to implement a spectrum allocation approach on the end node side than on a centralized unit. While taking into consideration that no extra processing can be afforded at the end node where every Watt is counted at transmission to save energy.

To be compatible with the constraint of low complexity of the end node hardware, a kind of Artificial Intelligence (AI) algorithm is considered. This proposed approach is based on reinforcement learning algorithms which have been introduced by the machine learning community [115] and first proposed for cognitive radio communications more than 10 years ago [116]. It is also experimentally validated on real radio signals for cognitive radio and especially for Opportunistic Spectrum Access (OSA) in [117]. As asserted for OSA, Multi-Armed Bandit (MAB) problem can be utilized to model the IoT spectrum access issue [113]. Thus, reinforcement learning is based on a feedback loop that gives a success|failure measure of the action. In the IoT context, a binary reward, i.e. 1|0 for the presence|absence of the ACK packet which is sent by the gateway to the end node, is considered, as shown in Figure 6.1. If a message has been successfully transmitted from the end node to the gateway on the uplink channel, as well as the ACK message has been successfully transmitted from the gateway and received by the end node on the same channel in the downlink, a reward of "1" is given plus the channel parameters, i.e. SNR, RSSI and ESP, are estimated to optimize and accelerate the learning process as illustrated in the following sections. While a reward of "0" means that packet loss has happened either on uplink or downlink so that ACK has not been received by the end node. Maximizing the PDR is the main target, or equivalently, maximizing its cumulated reward. This proposed approach can adapt to any other IoT protocol, moreover, it has the following main advantages:

- Coordination between the end nodes is not necessary. Therefore, no extra retransmission, no data to be added into the uplink or downlink packet. While the content of the ACK packet isn't changed.
- The very low processing and memory overhead of both the implementation and execution of the proposed approach [113]. Consequently, it is possible to execute the proposed algorithm in the end node devices whose complexity is negligible in terms of processing, hardware, memory as well as energy consumption overhead.

- The strong mathematical proof of convergence of the Bandit algorithms [118]. Thanks to the good matching between models and reality, these proofs are verified in real radio conditions [117].
- Rapid convergence of the proposed learning techniques in real experiments [119].
- Thanks to the reinforcement learning concept, any prior training isn't needed. This proposed algorithm can efficiently start learning from scratch.
- These proposed learning algorithms' results will always outperform the given state-of-the-art ones with random frequency allocation [119].

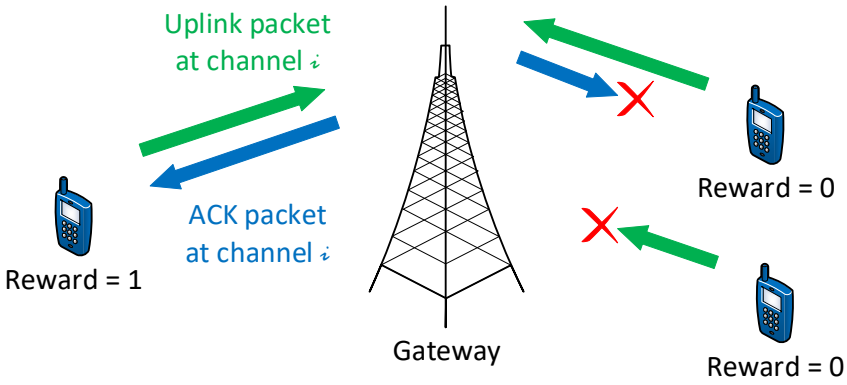


Figure 6.1 – An end node sends a packet with the acknowledged messaging mode, whereas the reward value depends on the presence|absence of the ACK packet.

6.6 Proposed Reinforcement Learning Techniques

Bandit algorithms are used at the end node side to handle IoT wireless spectrum issues. While LoRaWAN is the considered example, a simple ALOHA-based protocol is the utilized communications between the end nodes and the gateway. Whenever the end nodes decide, they can transmit their packets in one of the $K \geq 2$ channels which are predefined in frequency f_i. As stated in the previous sections, the channel conditions in unlicensed ISM bands suffer in particular from propagation conditions and interference which are different from one place to another. Even in one place, these channel conditions are very often unevenly distributed over the K different channels.

From the point of view of a single end node in the network, it has to choose one channel (or arm), denoted as $i \in \{1, ..., K\}$, every time slot n when it sends a packet to

the gateway, then, it starts to wait for a fixed delay (one second in LoRaWAN) in the same channel i for receiving an ACK packet from the gateway, as shown in Figure 6.1. But due to propagation issues, the message sent by the end node to the gateway, or the ACK sent by the gateway to the end node, could be lost. Hence, selecting the channel i at time n yields a random feedback, i.e. the binary reward $r_i(n)$ and the channel quality $g_i(n)$. Maximizing the transmission success rate of the end node or, equivalently, maximizing the cumulative reward is the main target. Sequence of rewards drawn from a given arm i is assumed to be *i.i.d.*, consequently, this problem is considered as a "stochastic" MAB [118]. A player (here, an end node) has to try all arms (here, channels), a sufficient number of times to get a robust estimate of their qualities, while not selecting the worst arms too much. This requires tackling the so-called exploration-exploitation dilemma to be able to progressively focus on the best arm i.e., the arm with the largest average reward, using the proposed algorithms in the following subsections.

6.6.1 UCB Algorithm

Exploiting the channel with the highest estimated mean by selecting it at each time could be a first naive solution, however, this "greedy" approach is known to fail dramatically [115]. Thus, the selection of arms is highly dependent on the first draws with this greedy policy. For example, if the first transmission in one channel succeeds and the first one on the other channel fails, the end node will never use the other channel again, even it is the best one and has the best channel conditions on average. While UCB algorithm instead is adding an extra exploration term to the empirical mean, which can be viewed as a confidence bound [118]. At each end node, the number of times the channel i is selected up-to time index $n \geq 1$ is calculated as:

$$T_i(n) = \sum_{m=1}^{n} \mathbb{1}_{A(m)=i} \; \forall i, \tag{6.1}$$

where $A(m) \in \{1, ..., K\}$ is a discrete action which corresponds to the chosen channel index at time index m. Hence, the empirical mean estimation of the successful transmissions obtained in channel i by selecting it up to time n, is denoted by the mean reward as:

$$R_i(n) = \frac{1}{T_i(n)} \sum_{m=1}^{n} r_i(m) \mathbb{1}_{A(m)=i} \; \forall i. \tag{6.2}$$

Subsequently, the upper confidence bound $B_i(n)$ of the channel i is denoted by the sum of the empirical mean and a confidence term as:

$$B_i(n) = R_i(n) + \alpha\sqrt{\frac{\ln n}{T_i(n)}}, \tag{6.3}$$

where the exploration factor α is recommended by the theory to be ≥ 0.25 [120]. At this point, the maximum upper confidence bound $B_i(n)$ is chosen by the end node to decide the next channel for sending the packet, thus, the next action $A(n+1)$ is obtained by choosing the best channel as:

$$A(n+1) = \arg\max_i(B_i(n)). \tag{6.4}$$

For each end node, the time index n corresponds to the total number of transmitted packets from the beginning. As time is not slotted, this time index n isn't shared across different end nodes. Every end node implements its own UCB algorithm independently, which can be described as follows in Algorithm 1.

Data: $K, [\alpha]$
Result: $A(n+1)$
for $n = 1$ *to* ∞ **do**
  if $n < K$ **then**
    Initialize policy by trying each channel for at least one time.
    $A(n+1) = n+1$
  else
    $T_i(n) = \sum_{m=1}^{n} \mathbb{1}_{A(m)=i}\ \forall i$
    $R_i(n) = \frac{1}{T_i(n)} \sum_{m=1}^{n} r_i(m)\mathbb{1}_{A(m)=i}\ \forall i$
    $B_i(n) = R_i(n) + \alpha\sqrt{\frac{\ln n}{T_i(n)}}$
    $A(n+1) = \arg\max_i(B_i(n))$
  end
end

Algorithm 1: UCB policy as depicted in [118].

6.6.2 QoC-A Algorithm

As mentioned in the previous sections, there is a strong dependency between the PDR and the channel quality. Hence, a QoC-A policy is proposed to find the channel which is optimal in terms of both quality and probability of receiving the ACK packet [121]. As first stated in Algorithm 2, all the channels are explored at least once to acquire the initial binary reward and their channel qualities. At the final step, Algorithm 2 returns the channel index $A(n+1)$ which has to be used in the next time slot. Again, $T_i(n)$ defines the number of times channel i has been utilized up to time n. After $n \geq K$ iterations, the term $B_i(n)$, i.e. corresponding to the score of the $i-$th channel at time n, are updated as [121]:

$$B_i(n) = R_i(n) - Q_i(n) + \alpha\sqrt{\frac{\ln n}{T_i(n)}}, \tag{6.5}$$

where $R_i(n)$ denotes the exploitation term, or equivalently the empirical mean of the states (ACK packet received or not) of the $i-$th channel at time n. While the bias term $\alpha\sqrt{\frac{\ln n}{T_i(n)}}$ forces to explore the other channels, if the scheme leads to a channel whose average rewards degrades. Furthermore, the quality term $Q_i(n)$ defines the quality information of the channel i which can be calculated as:

$$Q_i(n) = \beta(1 - \frac{G_i(n)}{G_{max}(n)})\frac{\ln n}{T_i(n)} \tag{6.6}$$

with

$$G_i(n) = \frac{1}{T_i(n)} \sum_{k=1}^{T_i(n)} g_i(k) \tag{6.7}$$

and

$$G_{max}(n) = \max_i G_i(n), \tag{6.8}$$

where $G_i(n)$ is the empirical mean of quality observations $g_i(n)$ collected from channel i and $G_{max}(n)$ is the maximum expected quality value within the set of channels. Moreover, the new parameter β forces the algorithm to give some weight to the quality in the score computation, whereas the parameter α forces the exploration of other channels to check their probability of receiving the ACK packet. Such a formulation tends to select a channel with the highest quality and probability to be acknowledged.

An important contribution of this chapter compared to the previous works in the cognitive radio field [121], is that the quality observation $g_i(n)$ is the channel quality,

i.e. ESP in linear scale, of the ACK packet. On the contrary, the channel parameters are obtained from a sensing phase before sending the packet in the OSA context, as the "listen-before-talk" model in [122], which consumes power. It also worth mentioning that ESP is the used parameter that is obtained as a function of SNR, thus, it has an enlarged range when the SNR is very low, unlike the RSSI which has a limitation as detailed in Section 4.4.4. Accordingly, this proposed algorithm is simple to implement and to use in practice, even on embedded microprocessors with limited computation and memory capabilities as detailed in the following sections.

Data: $K, [\alpha, \beta]$
Result: $A(n+1)$
for $n = 1$ *to* ∞ **do**
  if $n < K$ **then**
    Initialize policy by trying each channel for at least one time.
    $A(n+1) = n+1$
  else
    $T_i(n) = \sum_{m=1}^{n} \mathbb{1}_{A(m)=i} \; \forall i$
    $R_i(n) = \frac{1}{T_i(n)} \sum_{m=1}^{n} r_i(m) \mathbb{1}_{A(m)=i} \; \forall i$
    $G_i(n) = \frac{1}{T_i(n)} \sum_{k=1}^{T_i(n)} g_i(k)$
    $G_{max}(n) = \max_i G_i(n)$
    $Q_i(n) = \beta(1 - \frac{G_i(n)}{G_{max}(n)}) \frac{\ln n}{T_i(n)}$
    $B_i(n) = R_i(n) - Q_i(n) + \alpha \sqrt{\frac{\ln n}{T_i(n)}}$
    $A(n+1) = \arg\max_i (B_i(n))$
  end
end

Algorithm 2: QoC-A policy.

6.6.3 DQoC-A Algorithm

In many IoT applications, the end node could be in a non-stationary scenario, for instance, when it moves across different locations whose CSI shapes are different. Based on that, the probability of receiving the ACK packet and the quality of each channel are likely to experience changes in time, which exhibits the limitation of the aforementioned MAB algorithms. Although the convergence of these algorithms is very fast, however, the stationarity of the environment is required. Resetting the learning algorithm from time to time can be a simple solution, however, it may fail to determine when the propagation

conditions change as the acquired reward distribution evolves with time occasionally.

In this chapter, a DQoC-A algorithm is proposed to learn on the same channel conditions, i.e. quality and the probability of receiving the ACK packet, as the previous QoC-A but in non-stationary scenarios [123]. Since the confidence interval of the standard QoC-A policy becomes tighter when time goes up, it is not appropriated for the non-stationary environment as stated before. While the motivation for the DQoC-A policy is to find an optimal channel in the case of changing environments, with less exploration. Therefore, discount factors (λ and λ_g) are considered for the DQoC-A to guaranty the adaptiveness of DQoC-A policy in a non-stationary environment, as stated in Algorithm 3. The idea behind the inclusion of these discount factors (λ and λ_g) is to give more weight to recent observations compared to the ones acquired in the past. A remarkable contribution of this chapter compared to the previous work in [123], is that two different discount factors (λ and λ_g) are considered rather than using one. Thus, λ and λ_g should have different values as they are used to average two different distributions, which are the binomial distribution of the binary reward $r_i(n)$ and the log-normal distribution of the channel quality $g_i(n)$, respectively.

Based on that, this proposed DQoC-A policy learns a channel that is optimal in terms of the probability of receiving the ACK packet and quality in a gradual manner. As with QoC-A policy, an end node employing DQoC-A policy first starts to explore all channels at least once initially. After $n \geq K$ iterations, it updates the scoring term $B_i(n)$, however, each term in the equation is adapted to take into account the non-stationary hypothesis as:

$$B_i(n) = R_i(n) - Q_i(n) + \alpha \sqrt{\frac{\ln W(n)}{N_i(n)}}, \tag{6.9}$$

where $N_i(n)$ is the discounted number of times channel i has been used up to time n, and $W(n)$ is the total discounted time. Contrary to QoC-A policy, the empirical mean of rewards $R_i(n)$ and channel quality $G_i(n)$ are estimated by taking into account the discount factors ($0 < \lambda < 1$ and $0 < \lambda_g < 1$), as shown in Algorithm 3. While the coefficients α and β, are the same as in the QoC-A policy, to weight exploration for the probability of receiving the ACK packet and channel quality, respectively.

Data: $K, [\alpha, \beta, \lambda, \lambda_g]$
Result: $A(n+1)$
for $n = 1$ *to* ∞ **do**
  if $n < K$ **then**
    Initialize policy by trying each channel for at least one time.
    $A(n+1) = n+1$
  else
    $N_i(n) = \sum_{m=1}^{n} \lambda^{n-m} \mathbb{1}_{A(m)=i} \; \forall i$
    $W(n) = \sum_{i=1}^{K} N_i(n)$
    $R_i(n) = \frac{1}{N_i(n)} \sum_{m=1}^{n} \lambda^{n-m} r_i(m) \mathbb{1}_{A(m)=i} \; \forall i$
    $Ng_i(n) = \sum_{m=1}^{n} \lambda_g^{n-m} \mathbb{1}_{A(m)=i} \; \forall i$
    $G_i(n) = \frac{1}{Ng_i(n)} \sum_{m=1}^{n} \lambda_g^{n-m} g_i(m) \mathbb{1}_{A(m)=i} \; \forall i$
    $G_{max}(n) = \max_i G_i(n)$
    $Q_i(n) = \beta(1 - \frac{G_i(n)}{G_{max}(n)}) \frac{\ln W(n)}{N_i(n)}$
    $B_i(n) = R_i(n) - Q_i(n) + \alpha \sqrt{\frac{\ln W(n)}{N_i(n)}}$
    $A(n+1) = \arg\max_i (B_i(n))$
  end
end

Algorithm 3: DQoC-A policy.

6.7 Experiment Setup

Comparing the proposed algorithms against the state-of-the-art method with random frequency allocation in real conditions of operation, is the main target of the experiment. This is done by setting the LoRaWAN configuration as presented in Section 4.4.3, but it could be done with any other IoT standard, as soon as it uses acknowledged messaging. Again, the Tektelic KONA Macro Gateway is used whose antenna is fixed on the roof of the university building [104], as shown in Figure 4.2a and 4.9a. While the end node is implemented using a Pycom card which is composed of an Expansion Board and a LoPy 4 module [105], as shown in Figure 6.2.

First, the network is joined with an OTAA. Then, the proposed algorithms are executed sequentially at each time index n inside the Pycom node. An uplink packet is transmitted at the channel i which is chosen by the algorithm, then the node waits about 10 seconds before sending the next packet to respect the duty cycle. By default, the gateway attempts to send one acknowledgment at the same channel i and center frequency f_i as the message transmitted. Subsequently, the node writes the information of the last

received downlink packet (packet number, ESP, etc.) to the payload of the next uplink packet. In this experiment, there is no retransmission attempt if the node does not receive the ACK packet. A desktop computer that runs a Python program is used as an AS only for analyzing purposes in this experiment. This computer receives data from the LNS, as well as LoRa metadata with all parameters of the LoRaWAN transmission (f_i, SF, W, RSSI, SNR, etc.). Those data are analyzed as detailed in the following sections. They are provided to the research community on this online repository [124].

Figure 6.2 – The packaged Pycom device is fixed on a rod and connected to a battery inside a building.

6.8 Experimental Results

The scope of this section is to compare the aforementioned learning policies against the state-of-the-art ones with random frequency allocation that is considered as the reference. For emulating the state-of-the-art frequency allocation, a uniform Round-Robin algorithm is executed that simply transmits the packets sequentially at each frequency band f_i across the time index n. Furthermore, the experiments are performed in two different environments as detailed in the following subsections.

6.8.1 Scenario 1: Stationary IoT node

The first experiment is carried out by placing the end node in a fixed position without moving it to have stationary channel conditions. Hence, the CSI shapes are almost identical across the whole experiment duration of $\approx$ 8.8 hours and 800 iterations (800 transmitted packets), as shown by the received power level in Figure 6.3. Each CSI shape is acquired by averaging the ESP values at each frequency band f_i independently every 200 iterations. Moreover, the frequency selectivity of this CSI is clear, particularly at 867.3 MHz with a deep fade of more than 10 dB depth. These unequal channel qualities across the frequency bands f_i could be exploited using the proposed algorithm QoC-A.

As depicted in Figure 6.4, the performance of UCB ($\alpha = 0.6$), and QoC-A ($\alpha = 0.6$ and $\beta = 0.2$), i.e. with ESP in the linear scale as the quality observation $g_i(n)$ that is shown in Algorithm 2, are compared against the uniform frequency allocation. As promised in the previous sections, all the learning algorithms outperform the uniform frequency allocation. On the other hand, the final average reward obtained using the QoC-A policy outperforms the UCB policy, as manifested in Figure 6.4a. Accordingly as shown in Figure 6.4b, the cumulative regret obtained for all the methods preserve the same performance rank over the whole time index n with final total lost packets of 32, 39 and 132, while using the proposed algorithm QoC-A, the classical UCB, and the uniform frequency allocation, respectively. As shown in Table 6.1, QoC-A policy outperforms all the other algorithms by a total number of successfully transmitted packets of 768 over 800. This indicates that the proposed policy QoC-A is exploiting properly the channel quality to converge faster. Hence, it does not need to lose some time to acquire this knowledge before learning actually ends.

Table 6.1 – Number of successfully transmitted and lost packets in the stationary scenario

Algorithm	**Succeed**	**Lost**
Uniform	668	132
UCB	761	39
QoC-A	768	32

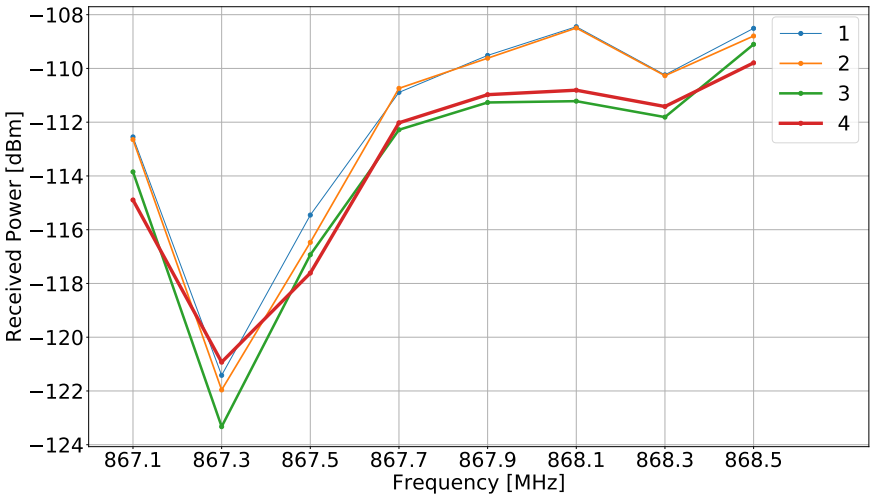
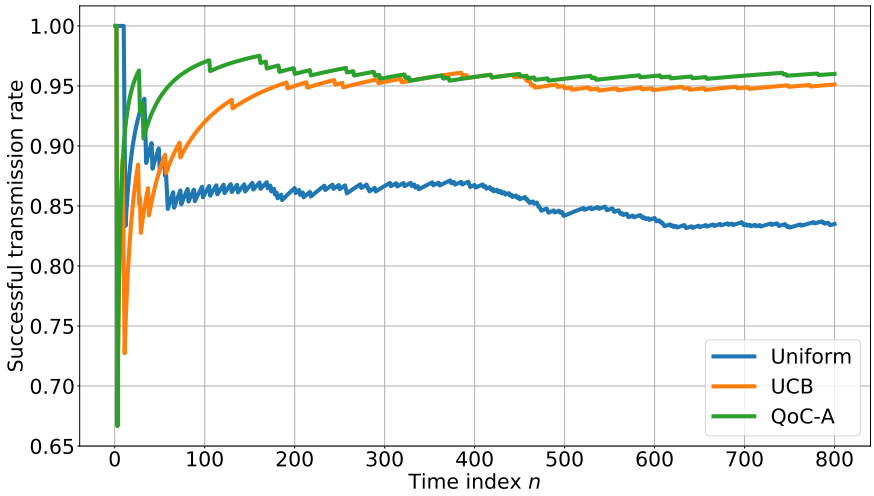


Figure 6.3 – Stability of the CSI shape throughout the 800 iterations (one CSI every 200 iterations) in the stationary scenario.

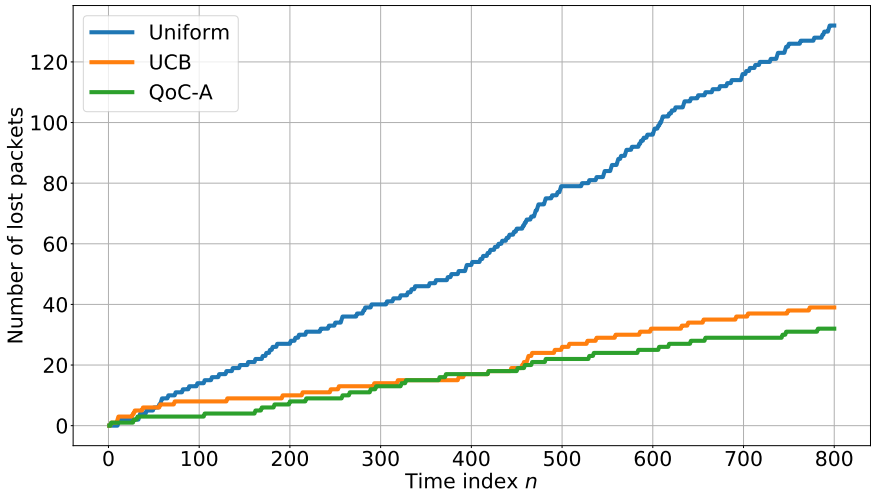
6.8.2 Scenario 2: Non-Stationary IoT node

For analyzing the behavior of MAB learning policies in a non-stationary scenario with an abrupt changing environment, the end node is moved during the experiment across three different positions to have unconstant channel conditions. As shown in Figure 6.5, the CSI shapes are different across the whole experiment duration of ≈ 6.6 hours. Again, each CSI shape is estimated by averaging the ESP values every 200 iterations at each frequency band f_i independently. Although all the locations are indoor, nevertheless, a progressive reduction is observed in the received power across locations 1 to 3. This reduction is reasonable as the end node goes deeper inside the building from location 1, i.e. behind the window as shown in Figure 6.2, to location 3. Moreover, the CSI shape which is represented in location 1 shows fades of $\approx$5 dB, especially at 867.1 MHz and 868.5 MHz, while the best channel quality is obtained at 867.9 MHz. The opposite behavior is obvious in the CSI shapes of locations 2 and 3 whose best channel conditions are almost around 868.1 MHz, 868.3 MHz and 868.5 MHz. While they have degradation in the channel quality at 867.9 MHz, contrary to the CSI of location 1. These abrupt changes in the channel qualities from one location to another could be learned using the proposed algorithm DQoC-A.

In this non-stationary environment, identifying the abrupt change in the reward distribution with reduced delay should be the ability of an optimal policy. As shown in



(a) Average reward across the iteration number.



(b) Cumulative regret across the iteration number.

Figure 6.4 – Result of the algorithms against the uniform frequency allocation (non-learning) in the stationary scenario.

Figure 6.6, the performance of UCB ($\alpha = 0.6$), QoC-A ($\alpha = 0.6$ and $\beta = 0.2$) and DQoC-A ($\alpha = 0.6$, $\beta = 0.2$, $\lambda = 0.98$ and $\lambda_g = 0.90$) are compared against the naive frequency allocation. For both the QoC-A and DQoC-A algorithms, the linear scale of the ESP value is utilized as the quality observation $g_i(n)$. Moreover, the discount factor λ is equal to 0.98 in the DQoC-A policy, while λ_g is set to 0.90 to acquire more rapidly the most recent ESP values, which follows a log-normal distribution from the channel shadowing when the end node moves and subsequently to converge faster. Motions of the end node are represented by the two breakpoints at time index $n = 200$ and $n = 400$ in which an abrupt change in the reward distribution is indicated. Before the first breakpoint (at $n = 200$), the evolution of the cumulative regret is almost flat using any algorithm. This is a reasonable behavior as most of the frequency bands f_i at location 1 have relatively high ESP values and low packet losses. While after the first breakpoint (at $n = 200$), DQoC-A achieves significantly lower regret and higher average reward than QoC-A and UCB policy in this non-stationary environment. As shown in Table 6.2, DQoC-A policy outperforms all the other algorithms by a total number of successfully transmitted packets of 520 over 600. This is plausible due to the inclusion of the discount factors (λ and λ_g) in the DQoC-A calculation, so it wastes significantly less time than QoC-A to identify an abrupt change (at $n = 200$ or $n = 400$) in the channel conditions. While the UCB and QoC-A are prevented to adapt quickly to changes as the old rewards have a higher influence on them. In other words, all the learning algorithms converge to the optimal mean reward in the long run, but DQoC-A policy benefits from less dependency on past observations.

Table 6.2 – Number of successfully transmitted and lost packets in the non-stationary scenario

Algorithm	**Succeed**	**Lost**
Uniform	405	195
UCB	487	113
QoC-A	497	103
DQoC-A	520	80

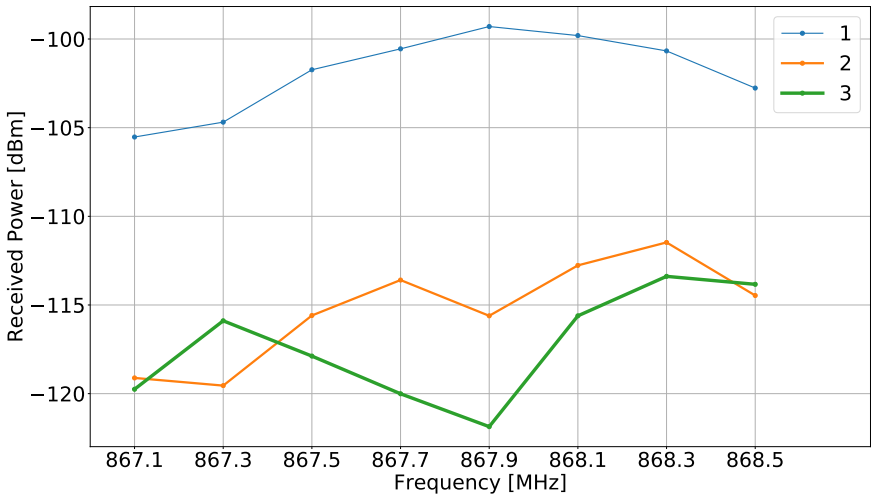
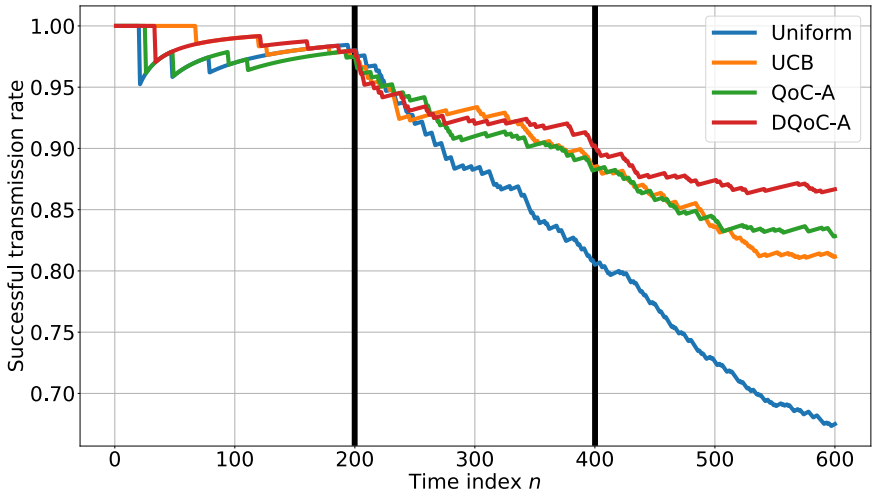


Figure 6.5 – Evolution of the CSI shape throughout the 600 iterations (one CSI every 200 iterations) in the non-stationary scenario.

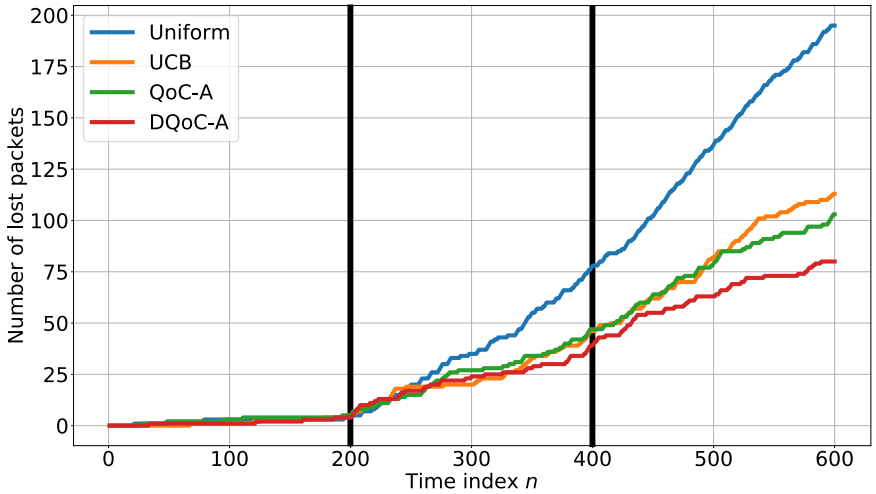
6.9 Additional Remarks on the Algorithms' Configurations

The aforementioned results confirm the feasibility of utilizing the proposed policies to achieve lower packet losses in different scenarios. However, the parameters of all the proposed learning algorithms should be adjusted based on the environment's conditions. Thus, the exploration factor α is chosen depending on the channel conditions in terms of channel qualities or presence of interference signals [125]. Furthermore, the influence of the optimal choice of α is higher as the number K of channels increases, and less as it decreases. In the case of a small number of arms, choosing the value of the exploration factor α does not influence the policy performance much. In the opposite case, if the number of arms to explore is large, the choice of this factor will have a greater influence. If, for example, the chosen alpha parameter is deviated from the optimal value for a large number of arms to be explored, significant degradation in performances may be observed. Hence, the algorithm will then have either a tendency either to over-explore or to over-exploit.

On the other hand, the discount factors (λ and λ_g) values should be adjusted based on the requirements of the potential application. Thus, they will regulate properly the amount of dependency on the old observations to track an abrupt change in the reward



(a) Average reward across the iteration number.



(b) Cumulative regret across the iteration number.

Figure 6.6 – Result of the algorithms against the uniform frequency allocation (non-learning) in the non-stationary scenario.

distribution. For example, if an IoT node sends one packet every a fixed short time, therefore, the discount factors in the DQoC-A algorithm should be close to one to benefit also from the older observations as the environment is stationary over such a short horizon of time, and vice versa.

6.10 Summary and Conclusion

This chapter introduces a decentralized learning technique to mitigate the channel impairments in the IoT signal propagation. Thus, the proposed QoC-A policy could learn a proper frequency allocation based on the channel quality, i.e. ESP, as an extra observation to avert using the channels whose quality is poor. Besides, another DQoC-A policy is proposed to adapt rapidly to any change in the channels' conditions, as a result of the IoT node motion for example. To demonstrate the low complexity of these proposed algorithms and the feasibility of implementing them on the IoT device side, at a very low cost of implementation and no protocol overhead, a real experiment campaign is carried out in the city of Rennes. Consequently, the results show that QoC-A has a more optimized and accelerated learning process than the classical UCB policy, and then it has lower packet losses at the end of the process. On the other hand, DQoC-A policy converges faster than QoC-A policy in the non-stationary scenario, thanks to the discount factor in DQoC-A algorithm which decreases the dependency on the old observations gradually.

For future work, these proposed reinforcement learning algorithms could be implemented to decrease the packet losses in the potential IoT applications. Thus, the configurations of the implemented policy should be adjusted based on the application.

Chapter 7

General Conclusions and Future Works

7.1 Conclusions

In this thesis, the propagation properties of the wireless channel are utilized for different purposes which all have an impact on the overall efficiency of the IoT functions, especially in LPWAN. Thus, the thesis is based on the following two axes:

1. **Localization** is optimized after presenting background on the passive localization techniques which are implemented on the gateway side with low energy consumptions. Thus, strategies and methods are proposed to improve the accuracy of the most popular LPWAN localization technique, i.e. TDoA. To obtain a more precise localization, these proposed approaches could be used as independent or complementary methods. Contrary to the classical TDoA technique which localizes very often on the space far away from the vicinity of node location, a parametric TDoA technique is proposed to localize the node on one of the hyperbolas. Over different timestamps drift variances and number of gateways, this proposed parametric method outperforms the classical one, therefore, this method is asserted to be less sensitive to any perturbation in the timestamps or the TDoA values.

   A preprocessing algorithm to drop out the perturbated timestamps before localizing using any TDoA technique is proposed. For implementing this approach, an initial guess of the node location and the euclidean distances from it to each hyperbola are the only requirement. After examining the most perturbed hyperbolas whose locations are far away from the vicinity of the node location, the perturbated timestamps are detected. After dropping these outlier timestamps, the localization accuracy is improved, as any unreliable sensitivity from them is prevented.

By utilizing a paired CRLB as an instrumental tool, another preprocessing method is proposed to detect the outlier TDoA values. The only requirement for this method is an initial guess of the node location, hence, the paired CRLB of each TDoA combination is estimated from the vicinity of this initial guess location. Based on a robust thresholding technique, the low paired CRLB are detected and then their corresponding TDoA combinations are dropped. By localizing with the best TDoA values whose perturbation probabilities are less, this proposed method is proven to decrease the localization error.

On the other hand, the feasibility of localizing using LoRa CSI is investigated by performing outdoor measurement campaigns in the area of Beaulieu Campus in Rennes. In contrary to the RSSI values, the spatial and temporal stability of CSI are asserted to be relatively stronger. Using the PCC, the CSI reciprocity is quantified which shows a very high linear correlation between the uplink and downlink CSI, at each different location. On the contrary, most of the CSI shapes from different locations are highly uncorrelated to each other. By utilizing the frequency hopping in the LoRa systems to get access to a wider band, a significant localization gain could be achieved. Moreover, the reliability of using the ESP is proven by its enlarged range when the SNR is very low, unlike the RSSI which has a limitation. Based on that, a wider CSI can be obtained by the frequency hopping after estimating the narrowband CSI from the IQ samples of the received LoRa preamble chirps at each center frequency. Or, by averaging the ESP values for each center frequency until the wider CSI shape is acquired by integrating these averaged values.

2. **Spectrum Allocation** algorithms are proposed for the IoT, after studying the frequency dependency of PDR. During a measurement campaign in the city of Rennes, the channel parameters (RSSI, ESP, SINR) from different node locations are estimated. Hence, PDR is investigated against ESP at each different frequency. At each different location, the PDR values is proven to be correlated with the CSI shape. On the other hand, the presented data indicate that PDR has a more natural relationship with ESP than RSSI. Based on these investigations, an ESP parameterized beta distribution function is deployed to model PDR. For the LoRaWAN

network regulation and optimization, this proposed model could be utilized.

The aforementioned frequency dependency of ESP is exploited to improve PDR. Hence, a decentralized learning technique is proposed to mitigate the channel impairments in the IoT signal propagation. A proposed QoC-A policy considers the ESP value as an extra observation of the channel quality information to learn a proper frequency allocation. Accordingly, this proposed algorithm averts using the channels whose quality is poor. Moreover, to adapt rapidly to any change in the channels' conditions, as a result of the IoT node motion, for example, another DQoC-A policy is proposed. An experimental campaign is carried out in the city of Rennes, hence, the low complexity and the feasibility of implementing these proposed algorithms on the IoT device side with no protocol overhead are demonstrated. With the obtained results at the end of the process, the QoC-A policy is proven to have lower packet losses, thus, its learning process is more optimized and accelerated than the classical UCB policy. While in the non-stationary scenario, DQoC-A policy converges faster than QoC-A policy as a result of the discount factor inclusion in the DQoC-A algorithm, which decreases the dependency on the old observations gradually.

7.2 Future Works

In this thesis, the presented works in each contribution suggest many possible directions for future studies, such as the following perspectives:

- The parametric TDoA can be optimized, for instance, by merging its two main steps.
- For outlier timestamps and TDoA values detection, an optimum weighting technique could be investigated rather than using a thresholding method.
- Machine learning techniques could be used to integrate other signal measurements with the proposed TDoA techniques such as RSSI or CSI.
- Utilizing ESP for all the potential IoT applications is highly recommended, especially localization using ESP fingerprinting instead of using RSSI in LoRaWAN.
- The proposed PDR model can be utilized to have an estimation of the PDR against ESP for having a proper envision of the LoRaWAN network design.
- To decrease the packet losses in IoT, the proposed spectrum allocation algorithms

could be optimized or implemented properly. Therefore, many hints are presented for adjusting properly the configurations of the proposed learning algorithms based on the application requirements.

BIBLIOGRAPHY

[1] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios," *IEEE Wireless Communications*, vol. 23, pp. 60–67, October 2016.

[2] "Lora alliance homepage." `https://lora-alliance.org`, note = Accessed: 2021-04-30,.

[3] R. Oliveira, L. Guardalben, and S. Sargento, "Long range communications in urban and rural environments," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 810–817, July 2017.

[4] B. C. Fargas and M. N. Petersen, "Gps-free geolocation using lora in low-power wans," in *2017 Global Internet of Things Summit (GIoTS)*, pp. 1–6, June 2017.

[5] C. Gu, L. Jiang, and R. Tan, "Lora-based localization: Opportunities and challenges," 2019.

[6] 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, pp. 34–40, Sep. 2017.

[7] L. Mainetti, L. Patrono, A. Secco, and I. Sergi, "An iot-aware aal system for elderly people," in *2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, pp. 1–6, July 2016.

[8] A. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, pp. 24–40, July 2005.

[9] Z. Li, T. Braun, X. Zhao, Z. Zhao, F. Hu, and H. Liang, "A narrow-band indoor positioning system by fusing time and received signal strength via ensemble learning," *IEEE Access*, vol. 6, pp. 9936–9950, 2018.

[10] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," in *Proceedings of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, WiNTECH '18, (New York, NY, USA), p. 89–97, Association for Computing Machinery, 2018.

[11] A. Abdelghany, B. Uguen, C. Moy, and D. Lemur, "On superior reliability of effective signal power versus rssi in lorawan," in *2021 28th International Conference on Telecommunications (ICT)*, pp. 1–5, June 2021.

[12] A. Abdelghany, B. Uguen, C. Moy, and D. Lemur, "Modelling of the packet delivery rate in an actual lorawan network," *Electronics Letters*, vol. 57, no. 11, pp. 460–462, 2021.

[13] Q. Liu, Y. Mu, J. Zhao, J. Feng, and B. Wang, "Characterizing packet loss in city-scale lorawan deployment: Analysis and implications," in *2020 IFIP Networking Conference (Networking)*, pp. 704–712, June 2020.

[14] S.-Y. Wang, Y.-R. Chen, T.-Y. Chen, C.-H. Chang, Y.-H. Cheng, C.-C. Hsu, and Y.-B. Lin, "Performance of lora-based iot applications on campus," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6, Sep. 2017.

[15] 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*, vol. 16, no. 9, 2016.

[16] G. Fettweis and E. Zimmermann, "Ict energy consumption - trends and challenges," *proc of WPMC*, vol. 4, 01 2008.

[17] M. Ajmone Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *2009 IEEE International Conference on Communications Workshops*, pp. 1–5, June 2009.

[18] G. Xu and Y. Xu, *GPS: Theory, Algorithms and Applications*. Springer Berlin Heidelberg, 2018.

[19] U. Noreen, A. Bounceur, and L. Clavier, "A study of lora low power and wide area network technology," in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, May 2017.

[20] P. Masek, M. Stusek , E. Svertoka , J. Pospisil, R. Burget, E. S. Lohan, I. Marghescu , J. Hosek, and A. Ometov, "Measurements of lorawan technology in urban scenarios: A data descriptor," *Data*, vol. 6, no. 6, 2021.

[21] Q. Guo, F. Yang, and J. Wei, "Experimental evaluation of the packet reception performance of lora," *Sensors*, vol. 21, no. 4, 2021.

[22] P. Robyns, P. Quax, W. Lamotte, and W. Thenaers, "A multi-channel software decoder for the lora modulation scheme," pp. 41–51, 01 2018.

[23] H. Kwasme and S. Ekin, "Rssi-based localization using lorawan technology," *IEEE Access*, vol. 7, pp. 99856–99866, 2019.

[24] A. Shojaifar, "Evaluation and improvement of the rssi-based localization algorithm : Received signal strength indication (rssi)," Master's thesis, Department of Software Engineering, 2015.

[25] 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*, MSWiM '16, (New York, NY, USA), p. 59–67, Association for Computing Machinery, 2016.

[26] S. Sadowski and P. Spachos, "Rssi-based indoor localization with the internet of things," *IEEE Access*, vol. 6, pp. 30149–30161, 2018.

[27] M. Anjum, M. A. Khan, S. Ali Hassan, A. Mahmood, and M. Gidlund, "Analysis of rssi fingerprinting in lora networks," in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, pp. 1178–1183, June 2019.

[28] H. Sallouha, A. Chiumento, and S. Pollin, "Localization in long-range ultra narrow band iot networks using rssi," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[29] S. Seidel and T. Rappaport, "914 mhz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE Transactions on Antennas and Propagation*, vol. 40, pp. 207–217, Feb 1992.

[30] B. R. Stojkoska, I. N. Kosović, and T. Jagušt, "How much can we trust rssi for the iot indoor location-based services?," in *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pp. 1–6, Sep. 2017.

[31] Y. Li, Z. He, Y. Li, H. Xu, L. Pei, and Y. Zhang, "Towards location enhanced iot: Characterization of lora signal for wide area localization," in *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, pp. 1–7, March 2018.

[32] Y. Li, Y. Zhuang, X. Hu, Z. Gao, J. Hu, L. Chen, Z. He, L. Pei, K. Chen, M. Wang, X. Niu, R. Chen, J. Thompson, F. M. Ghannouchi, and N. El-Sheimy, "Toward location-enabled iot (le-iot): Iot positioning techniques, error sources, and error mitigation," *IEEE Internet of Things Journal*, vol. 8, pp. 4035–4062, March 2021.

[33] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proceedings of the IEEE*, vol. 97, pp. 404–426, Feb 2009.

[34] J. Shen, A. F. Molisch, and J. Salmi, "Accurate passive location estimation using toa measurements," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 2182–2192, June 2012.

[35] S. Leugner, M. Pelka, and H. Hellbrück, "Comparison of wired and wireless synchronization with clock drift compensation suited for u-tdoa localization," in *2016 13th Workshop on Positioning, Navigation and Communications (WPNC)*, pp. 1–4, Oct 2016.

[36] R. Kaune, J. Hörst, and W. Koch, "Accuracy analysis for tdoa localization in sensor networks," in *14th International Conference on Information Fusion*, pp. 1–8, July 2011.

[37] K. Krizman, T. Biedka, and T. Rappaport, "Wireless position location: fundamentals, implementation strategies, and sources of error," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, vol. 2, pp. 919–923 vol.2, May 1997.

[38] P. Müller, H. Stoll, L. Sarperi, and C. Schüpbach, "Outdoor ranging and positioning based on lora modulation," in *2021 International Conference on Localization and GNSS (ICL-GNSS)*, pp. 1–6, June 2021.

[39] H. Xiong, Z. Chen, B. Yang, and R. Ni, "Tdoa localization algorithm with compensation of clock offset for wireless sensor networks," *China Communications*, vol. 12, pp. 193–201, Oct 2015.

[40] Z. Li, D. C. Dimitrova, D. H. Raluy, and T. Braun, "Tdoa for narrow-band signal with low sampling rate and imperfect synchronization," in *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)*, pp. 1–8, May 2014.

[41] A. Badawy, T. Khattab, D. Trinchero, T. E. Fouly, and A. Mohamed, "A simple aoa estimation scheme," 2014.

[42] S. Monfared, T.-H. Nguyen, L. Petrillo, P. De Doncker, and F. Horlin, "Experimental demonstration of ble transmitter positioning based on aoa estimation," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 856–859, Sep. 2018.

[43] Y. Wang and K. C. Ho, "An asymptotically efficient estimator in closed-form for 3-d aoa localization using a sensor network," *IEEE Transactions on Wireless Communications*, vol. 14, pp. 6524–6535, Dec 2015.

[44] N. BniLam, D. Joosens, J. Steckel, and M. Weyn, "Low cost aoa unit for iot applications," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, March 2019.

[45] L. Bai, C.-Y. Peng, and S. Biswas, "Association of doa estimation from two ulas," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, pp. 1094–1101, June 2008.

[46] Q. Spencer, B. Jeffs, M. Jensen, and A. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 347–360, March 2000.

[47] R. Klukas and M. Fattouche, "Line-of-sight angle of arrival estimation in the outdoor multipath environment," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 342–351, Feb 1998.

[48] N. BniLam, D. Joosens, M. Aernouts, J. Steckel, and M. Weyn, "Loray: Aoa estimation system for long range communication networks," *IEEE Transactions on Wireless Communications*, vol. 20, pp. 2005–2018, March 2021.

[49] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "Csi-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, pp. 1300–1309, July 2013.

[50] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, 2017.

[51] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "Csi-mimo: Indoor wi-fi fingerprinting system," in *39th Annual IEEE Conference on Local Computer Networks*, pp. 202–209, Sep. 2014.

[52] X. Wang, L. Gao, S. Mao, and S. Pandey, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 763–776, Jan 2017.

[53] C. Ma, M. Yang, Y. Jin, K. Wu, and J. Yan, "A new indoor localization algorithm using received signal strength indicator measurements and statistical feature of the channel state information," in *2019 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5, Aug 2019.

[54] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with csi images," *IEEE Transactions on Network Science and Engineering*, vol. 7, pp. 316–327, Jan 2020.

[55] F. Wanderley, A. Misra, and V. Singh, "Localization techniques in the future iot: A review," in *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, pp. 57–64, July 2019.

[56] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen, "Robust tracking in cellular networks using hmm filters and cell-id measurements," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 1016–1024, March 2011.

[57] M. Bouet and A. L. dos Santos, "Rfid tags: Positioning principles and localization techniques," in *2008 1st IFIP Wireless Days*, pp. 1–5, Nov 2008.

[58] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, pp. 1067–1080, Nov 2007.

[59] D. Plets, N. Podevijn, J. Trogh, L. Martens, and W. Joseph, "Experimental performance evaluation of outdoor tdoa and rss positioning in a public lora network," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, Sep. 2018.

[60] I. Daramouskas, V. Kapoulas, and T. Pegiazis, "A survey of methods for location estimation on low power wide area networks," in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–4, July 2019.

[61] A. Mackey and P. Spachos, "Lora-based localization system for emergency services in gps-less environments," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 939–944, April 2019.

[62] F. U. Khan, M. Awais, M. B. Rasheed, B. Masood, and Y. Ghadi, "A comparison of wireless standards in iot for indoor localization using lopy," *IEEE Access*, vol. 9, pp. 65925–65933, 2021.

[63] C. Sanchez, B. Arpi, A. Vazquez-Rodas, F. Astudillo-Salinas, and L. I. Minchala, "Performance evaluation of rssi-based positioning system with low-cost lora devices," in *Proceedings of the 16th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor & Ubiquitous Networks*, PE-WASUN '19, (New York, NY, USA), p. 37–44, Association for Computing Machinery, 2019.

[64] J. Pospisil, R. Fujdiak, and K. Mikhaylov, "Investigation of the performance of tdoa-based localization over lorawan in theory and practice," *Sensors*, vol. 20, no. 19, 2020.

[65] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 6, pp. VI–553, April 2003.

[66] A. Bensky, *Wireless Positioning Technologies and Applications*. USA: Artech House, Inc., 2007.

[67] M. Marcelić, B. Sandrić, J. Jelenić, and M. Jurčević, "Determining location in lp-wan using multilateration," in *2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET)*, pp. 1–4, April 2019.

[68] D. Bissett, "Analysing tdoa localisation in lora networks," 2018.

[69] N. Podevijn, D. Plets, J. Trogh, L. Martens, P. Suanet, K. Hendrikse, and W. Joseph, "Tdoa-based outdoor positioning with tracking algorithm in a public lora network," *Wireless Communications and Mobile Computing*, vol. 2018, p. 1864209, May 2018.

[70] H. Sallouha, A. Chiumento, S. Rajendran, and S. Pollin, "Localization in ultra narrow band iot networks: Design guidelines and tradeoffs," *IEEE Internet of Things Journal*, vol. 6, pp. 9375–9385, Dec 2019.

[71] A. Pimpinella, A. E. C. Redondi, M. Nicoli, and M. Cesana, "Machine learning based localization of lorawan devices via inter-technology knowledge transfer," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, June 2020.

[72] M. Anjum, M. A. Khan, S. A. Hassan, A. Mahmood, H. K. Qureshi, and M. Gidlund, "Rssi fingerprinting-based localization using machine learning in lora networks," *IEEE Internet of Things Magazine*, vol. 3, pp. 53–59, December 2020.

[73] Y. Li, Z. He, Z. Gao, Y. Zhuang, C. Shi, and N. El-Sheimy, "Toward robust crowdsourcing-based localization: A fingerprinting accuracy indicator enhanced wireless/magnetic/inertial integration approach," *IEEE Internet of Things Journal*, vol. 6, pp. 3585–3600, April 2019.

[74] Z. Chen, M. I. AlHajri, M. Wu, N. T. Ali, and R. M. Shubair, "A novel real-time deep learning approach for indoor localization based on rf environment identification," *IEEE Sensors Letters*, vol. 4, pp. 1–4, June 2020.

[75] J. Purohit, X. Wang, S. Mao, X. Sun, and C. Yang, "Fingerprinting-based indoor and outdoor localization with lora and deep learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Dec 2020.

[76] F. Lemic, V. Handziski, M. Aernouts, T. Janssen, R. Berkvens, A. Wolisz, and J. Famaey, "Regression-based estimation of individual errors in fingerprinting localization," *IEEE Access*, vol. 7, pp. 33652–33664, 2019.

[77] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, F. Riaz, and S. J. Kwon, "Outlier detection in indoor localization and internet of things (iot) using machine learning," *Journal of Communications and Networks*, vol. 22, pp. 236–243, June 2020.

[78] J. Yoo, "Change detection of rssi fingerprint pattern for indoor positioning system," *IEEE Sensors Journal*, vol. 20, pp. 2608–2615, March 2020.

[79] C. Laoudias, R. Piché, and C. G. Panayiotou, "Device signal strength self-calibration using histograms," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, Nov 2012.

[80] P. Chen, J. Shang, and F. Gu, "Learning rssi feature via ranking model for wi-fi fingerprinting localization," *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 1695–1705, Feb 2020.

[81] T. Janssen, M. Weyn, and R. Berkvens, "A primer on real-world rss-based outdoor nb-iot localization," in *2020 International Conference on Localization and GNSS (ICL-GNSS)*, pp. 1–6, June 2020.

[82] W. Choi, Y.-S. Chang, Y. Jung, and J. Song, "Low-power lora signal-based outdoor positioning using fingerprint algorithm," *ISPRS International Journal of Geo-Information*, vol. 7, no. 11, 2018.

[83] Y. Lin, W. Dong, Y. Gao, and T. Gu, "Sateloc: A virtual fingerprinting approach to outdoor lora localization using satellite images," in *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 13–24, April 2020.

[84] M. Laaraiedh, L. Yu, S. Avrillon, and B. Uguen, "Comparison of hybrid localization schemes using rssi, toa, and tdoa," in *17th European Wireless 2011 - Sustainable Wireless Technologies*, pp. 1–5, April 2011.

[85] J. Zhang, G. Han, N. Sun, and L. Shu, "Path-loss-based fingerprint localization approach for location-based services in indoor environments," *IEEE Access*, vol. 5, pp. 13756–13769, 2017.

[86] N. S. Kodippili and D. Dias, "Integration of fingerprinting and trilateration techniques for improved indoor localization," in *2010 Seventh International Conference on Wireless and Optical Communications Networks - (WOCN)*, pp. 1–6, Sep. 2010.

[87] K.-J. Baik, S. Lee, and B.-J. Jang, "Hybrid rssi-aoa positioning system with single time-modulated array receiver for lora iot," in *2018 48th European Microwave Conference (EuMC)*, pp. 1133–1136, Sep. 2018.

[88] T. Sathyan, A. Sinha, and T. Kirubarajan, "Passive geolocation and tracking of an unknown number of emitters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, pp. 740–750, April 2006.

[89] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, pp. 2568–2599, thirdquarter 2019.

[90] D. W. Mitchell, "96.40 a property of hyperbolas and their asymptotes," *The Mathematical Gazette*, vol. 96, no. 536, p. 299–301, 2012.

[91] R. Bridson, "Fast poisson disk sampling in arbitrary dimensions," in *ACM SIGGRAPH 2007 Sketches*, SIGGRAPH '07, (New York, NY, USA), p. 22–es, Association for Computing Machinery, 2007.

[92] R. Kaune, "Accuracy studies for tdoa and toa localization," in *2012 15th International Conference on Information Fusion*, pp. 408–415, July 2012.

[93] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, pp. 61–80, July 2012.

[94] A. A. Ghany, B. Uguen, and D. Lemur, "A parametric tdoa technique in the iot localization context," in *2019 16th Workshop on Positioning, Navigation and Communications (WPNC)*, pp. 1–6, Oct 2019.

[95] Q. Song, S. Guo, X. Liu, and Y. Yang, "Csi amplitude fingerprinting-based nb-iot indoor localization," *IEEE Internet of Things Journal*, vol. 5, pp. 1494–1504, June 2018.

[96] L. Li, J. Ren, and Q. Zhu, "On the application of lora lpwan technology in sailing monitoring system," in *2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pp. 77–80, Feb 2017.

[97] M. R. Seye, B. Ngom, B. Gueye, and M. Diallo, "A study of lora coverage: Range evaluation and channel attenuation model," in *2018 1st International Conference on Smart Cities and Communities (SCCIC)*, pp. 1–4, July 2018.

[98] R. El Chall, S. Lahoud, and M. El Helou, "Lorawan network: Radio propagation models and performance evaluation in various environments in lebanon," *IEEE Internet of Things Journal*, vol. 6, pp. 2366–2378, April 2019.

[99] A. A. Ghany, B. Uguen, and D. Lemur, "A robustness comparison of measured narrowband csi vs rssi for iot localization," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pp. 1–5, Nov 2020.

[100] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Transactions on Information Theory*, vol. 54, pp. 2515–2534, June 2008.

[101] A. Rahmadhani and F. Kuipers, "Understanding collisions in a lorawan," *SURF Wiki*, 2017.

[102] R. Ghanaatian, O. Afisiadis, M. Cotting, and A. Burg, "Lora digital receiver analysis and implementation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1498–1502, May 2019.

[103] J.-J. van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Borjesson, "On channel estimation in ofdm systems," in *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, vol. 2, pp. 815–819 vol.2, July 1995.

[104] "Tektelic kona macro iot gateway. [online]. available:." `https://www.tektelic.com/uploads/Brochures/Kona%20Macro.pdf`.

[105] "Pycom documentation. [online]. available:." `https://GitHub.com/PyCom/PyCom-libraries`.

[106] "Measurement data. [online]. available:." `https://gitlab.com/ahmednagy/lorawan-beaulieu-measurement-2020.git`.

[107] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Pearson Correlation Coefficient*, pp. 1–4. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[108] 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, 2017.

[109] C. Moy, L. Besson, G. Delbarre, and L. Toutain, "Decentralized spectrum learning for radio collision mitigation in ultra-dense iot networks: Lorawan case study and experiments," *Annals of Telecommunications*, vol. 75, no. 11, pp. 711–727, 2020.

[110] J. Kerman, "A closed-form approximation for the median of the beta distribution," *arXiv preprint arXiv:1111.0433*, 2011.

[111] Y. Li, J. Yang, and J. Wang, "Dylora: Towards energy efficient dynamic lora transmission control," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pp. 2312–2320, July 2020.

[112] N. Chinchilla-Romero, J. Navarro-Ortiz, P. Muñoz, and P. Ameigeiras, "Collision avoidance resource allocation for lorawan," *Sensors*, vol. 21, no. 4, 2021.

[113] C. MOY, "Iotligent: First world-wide implementation of decentralized spectrum learning for iot wireless networks," in *2019 URSI Asia-Pacific Radio Science Conference (AP-RASC)*, pp. 1–4, March 2019.

[114] S. Rauh, T. Lauterbach, H. Lieske, J. Robert, and A. Heuberger, "Multipath characteristics of indoor-to-outdoor radio channels in the 868-mhz band," in *Smart SysTech 2016; European Conference on Smart Objects, Systems and Technologies*, pp. 1–11, July 2016.

[115] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[116] W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Upper confidence bound based decision making strategies and dynamic spectrum access," in *2010 IEEE International Conference on Communications*, pp. 1–5, May 2010.

[117] C. Moy, "Reinforcement learning real experiments for opportunistic spectrum access," in *WSR'14*, pp. 10–pages, 2014.

[118] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.

[119] L. Besson, R. Bonnefoi, and C. Moy, "Gnu radio implementation of malin: "multi-armed bandits learning for internet-of-things networks"," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, April 2019.

[120] S. Bubeck and C.-B. Nicolò. now, 2012.

[121] N. Modi, P. Mary, and C. Moy, "Qos driven channel selection algorithm for cognitive radio network: Multi-user multi-armed bandit approach," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, pp. 49–66, March 2017.

[122] N. Modi, P. Mary, and C. Moy, "Qos driven channel selection algorithm for opportunistic spectrum access," in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Dec 2015.

[123] N. Modi, *Machine learning and statistical decision making for green radio*. PhD thesis, CentraleSupélec, 2017.

[124] "Measurement data. [online]. available:." `https://gitlab.com/ahmednagy/lorawan-beaulieu-measurement-2021.git`.

[125] L. Melián-Gutierrez, N. Modi, C. Moy, F. Bader, I. Perez-Álvarez, and S. Zazo, "Hybrid ucb-hmm: A machine learning strategy for cognitive radio in hf band," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, pp. 347–358, Sep. 2015.

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

Titre : Exploitation ciblée du canal dans l'IoT : Amélioration de la localisation et de l'allocation du spectre dans les réseaux LPWAN

Mot clés : IoT, Localisation, PDR, ESP, Apprentissage automatique, Allocation de spectre

Résumé : L'objectif principal de ce travail est d'améliorer diverses fonctions LPWAN, en exploitant la connaissance des propriétés de propagation du canal sans fil. Tout d'abord, les techniques de localisation (sans GPS) sont étudiées, puis améliorées en proposant une technique TDoA paramétrique originale. Dans ce contexte, des techniques de prétraitement des observables spécifiques à la technique TDoA sont proposées. En utilisant un simulateur spécifique implémenté dans le cadre de la thèse, les résultats montrent que les approches proposées sont plus efficaces que la technique TDoA classique. D'autre part, l'utilisation du CSI pour la localisation est préconisée, la thèse étudie sa variabilité spatiale et temporelle. A travers une campagne de mesure, la fiabilité de l'utilisation du ESP en substitution du RSSI est affirmée par sa portée accrue. A ce stade, le LoRaWAN PDR est modélisé en fonction du ESP. Ensuite, le travail est orienté vers l'amélioration du PDR en proposant des algorithmes d'allocation de spectre qui exploitent la dépendance en fréquence du PDR. Ainsi, deux politiques décentralisées sont proposées pour apprendre un schéma d'allocation de fréquence approprié basé sur le ESP comme information sur la qualité du canal. Expérimentalement, les algorithmes proposés sont plus performants que la politique conventionnelle UCB avec moins de perte de paquets à la fin du processus.

Title: Purposely Exploiting the Channel in IoT: Enhancement of Localization and Spectrum Allocation in LPWAN

Keywords: IoT, Localization, PDR, ESP, Machine Learning, Spectrum Allocation

Abstract: The main objective of this work is to improve various LPWAN functions, by exploiting the knowledge of the propagation properties of the wireless channel. First, localization techniques (without GPS) are studied and then enhanced by proposing an original parametric TDoA technique. In this context, preprocessing techniques for observables specific to the TDoA technique are proposed. Using a specific simulator implemented in the framework of the thesis, the results show that the proposed approaches are more efficient than the classical TDoA technique. On the other hand, the use of CSI for localization is advocated, the thesis studies its spatial and temporal variability. Through a measurement campaign, the reliability of the use of ESP in substitution of RSSI is affirmed by its increased range. At this stage, the LoRaWAN PDR is modeled according to the ESP. Then, the work is oriented towards the improvement of the PDR by proposing spectrum allocation algorithms that exploit the frequency dependency of the PDR. Thus, two decentralized policies are proposed to learn an appropriate frequency allocation scheme based on the ESP as channel quality information. Experimentally, these proposed algorithms outperform the conventional UCB policy with less packet loss at the end of the process.