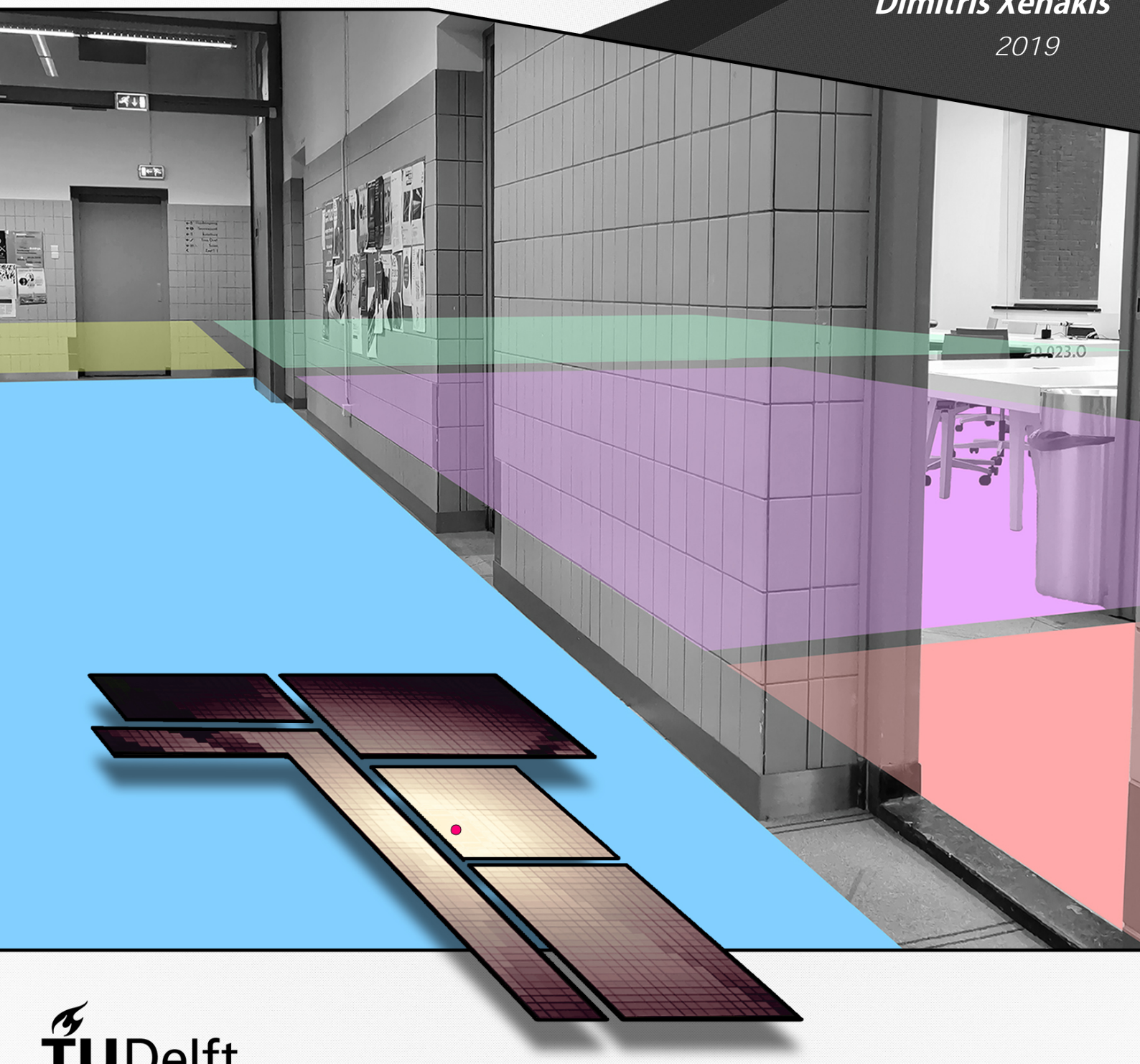
Placement optimization of Positioning Nodes:

Maximizing the distinction of Indoor Zones

MSc. Thesis
Geomatics for the Built Environment

Dimitris Xenakis

2019



TU Delft

COVER IMAGE:

Different zones within the Faculty of Architecture and the Built Environment at TU Delft. Being able to simulate (as further depicted) the propagation of the Bluetooth signals at any position, can enable the improvement of an Indoor Localization System.

Placement optimization of Positioning Nodes: Maximizing the distinction of Indoor Zones

MSc Thesis

by

Dimitris Xenakis

Supervisors:	Ir. E. Verbree	Department of GIS Technology
	Dr. ir. B.M. Meijers	Department of GIS Technology
Co-reader:	Dr. ir. A.A. (Sandra) Verhagen	Faculty of Civil Engineering and Geosciences
Ext. Supervisor:	S. Milani	Department of Architecture

Master of Science in Geomatics for the Built Environment

TU Delft

ABSTRACT

The performance of an Indoor Positioning System is highly related to the placement of the transmitting nodes that are used as references for the positioning estimations. Within this graduation project, we propose a methodology that can be used to optimize such a deployment and thus, increase the performance of an Indoor Positioning System that a) is based on Received Signal Strength Fingerprinting and b) is orientated towards providing location or zone estimations instead of exact positioning.

The optimization process involves 4 fundamental components. Firstly, the modeling of the obstructions in the indoor environment and also the zone modeling. Then, the definition of the performance metric that can be used to evaluate each different deployment scenario, in which case, our proposed metric considers the separation area and distances between the zones in the RSS vector space. The third component is the radio propagation model, required for simulating the transmitted signals from each node, where a model based on the ray tracing technique is selected. Finally, the last component is the selection of the optimization function that will control and drive the whole optimization process by choosing which deployment schemes to evaluate. For that, the utilization of a Genetic Algorithm has been selected.

The evaluation of our methodology showed that the most problematic regions in terms of localization accuracy are, as expected, those where different zones become adjacent. Yet, comparisons between regular node deployments and our optimized solutions indicated that, regardless the number of nodes, our optimization introduced in each case an overall localization improvement that was especially concentrated at the most problematic regions.

ACKNOWLEDGEMENTS

This document is the culmination of a long lasting research journey at TU Delft, upon the completion of which, I would like to take the opportunity to first express my heartfelt gratitude to both my mentors Edward and Martijn. To begin with, by delegating to me a subject which lies exactly in the epicenter of my interests has proved to be decisive for shaping the often laborious task of undertaking a graduation project, into a truly enjoyable experience. Moreover, their insights and advice have served as a lever in effectively orienteering my work and writing, steering me at all times towards the right track.

Specifically referring to Edward's influence on my work, a fact speaking for itself is that if it weren't for his course on location awareness, I would have probably never found interest in the field of indoor positioning and localization. Hence, not only this thesis, but also my forthcoming PhD pursuing on the subject of mobility prediction in indoor environments, would be non-existent.

Yet, I would additionally like to thank all those, family of course but also friends, who made my journey even easier without them even realizing it much. Either it is my fellow Geomatics students, or my beloved climbing group, or my Delft buddies, or my since-childhood friends; everyone helped in their own unique and deeply touching way.

Contents

List of Figures

Acronyms

AP Access Point. 2, 15

BLE Bluetooth Low Energy. vii, 1, 2, 4, 5, 7–11, 15, 19, 25, 26, 40

EM Electromagnetic. 1, 7

FSPL Free Space Path Loss. 40, 42

GA Genetic Algorithm. 6, 17, 18, 46, 47, 50, 73

GNSS Global Navigation Satellite Systems. 2, 17

GPS Global Positioning System. 9

IPS Indoor Positioning System. 1, 2, 8, 10, 11, 13, 21, 22, 53, 71, 73, 74

ISM Industrial, Scientific and Medical. 1, 7

kNN k-Nearest Neighbors. 10, 56, 68

LBS Location-Based Services. 2

RS Radio Signature. 22–25, 27, 56, 66, 68

RSSI Received Signal Strength Indicator. vii, viii, 9–13, 15, 16, 26, 27, 29, 40, 43, 45, 46, 51, 56, 74

SA Simulated Annealing. 18, 19

SNR Signal-to-Noise Ratio. 16

UWB Ultra-Wide Band. 9, 10

1

Introduction

Nowadays, it is rather easy to depict the importance of radio transmissions in every aspect of daily life, when finding a place on earth's surface having no artificial signals, is quite challenging. Even at the most distant places, where no cellular coverage, Television/Radio providers or other terrestrial broadcasting services exist and unless one is well hidden underground or underwater, various EM waves transmitted from space, will still be there.

Based on the application's characteristics and according to international regulations (International Telecommunication Union, 2016), the radio signals may propagate through reserved parts of the electromagnetic spectrum. For the case of industrial, scientific and medical applications, a corresponding Band (ISM) has been specified, where well-known wireless technologies such as Wi-Fi and Bluetooth are assigned (J.-S. Lee et al., 2007). A modified version of the latter, offering reduced power consumption while operating at the same frequency of 2.4GHz, namely the Bluetooth Low Energy (BLE), is the transmission technology that will be considered throughout this graduation project.

Since the development of the BLE specification, this technology has been constantly gaining attention in various fields, such as Health-care, Wellness and Sports, Home Automation, Internet of Things, etc. (Gomez et al., 2012). However, yet another BLE application of great importance would be Indoor Positioning and Navigation. This service typically suggests the installation of a mesh of BLE Beacons and a receiver that is able to "listen" to the transmitted signals. Then, depending on the positioning technique used, these signals are processed to finally produce an estimation of the receiver's position (Faragher and Harle, 2015).

As implied, an IPS can be based on various positioning techniques, such as triangulation, trilateration, multilateration, fingerprinting, etc. (H. Liu et al., 2007). The

fingerprint approach involves matching signal patterns with already known ones that have been georeferenced and stored within a database; while, it received this specific name, since these patterns are expected to be as unique, as a fingerprint can be. As noted in (Faragher and Harle, 2015), this method is the de-facto localization technique for (LBSs) on consumer devices today, while implementations using BLE signals have been shown to deliver among the most successful results (Faragher and Harle, 2015). Therefore, this combination (i.e. BLE-Fingerprinting) shall be respected throughout this graduation project.

1.1. Motivation and Problem Statement

Every mesh/network deployment could be evaluated based on some performance metric. Contrariwise, knowing this metric beforehand might enable the deployment of the network in such a way, that its performance becomes optimal. An example of such an optimization can be the distribution of cellular antennas in a town so that the coverage is maximized (Amzallag et al., 2005), or the deployment of Wi-Fi Access Points (APs) in a university so that the disconnections of walking users are minimized (Taufiq et al., 2011), or even the deployment of a GNSS constellation for maximum visibility (McKay and Pachter, 1997).

An IPS is a type of network that is straightforwardly coupled with the property of position. More than that, however, is also connected to the notion of location. Unlike a position, the description of which is narrowly directed by the numerical coordinates of a dimensionless point, a location can be perceived in many different ways. It can be all corridors, a desk, an open hall without physical boundaries, or even an elevator in a building. In all cases, the common characteristic is that the location can be considered as a superset of positions which are all attributed with the same thematic identifier; namely, the location where they belong.

Thus far, all research related to placement optimization of BLE nodes used in IPSs, has been done with respect to the positional accuracy. In practice, however, a BLE-based IPS often offers broader area/proximity estimations instead of positioning estimations of high accuracy. In these Location-enabled IPSs, the benefits of the so far proposed optimizations are not maximal, since the optimization process is highly consumed at enhancing aspects that have limited effect. Therefore, during this graduation project, a similar optimization problem will be faced. In this case, however, adjusting the placement of positioning nodes with the goal to increase the location prediction among different area sections (e.g. room A, B, etc.).

Besides the improvement in both the localization accuracy and navigation functionality, such an optimization could potentially increase also the cost-effectiveness ratio, leading ultimately to lower deployment costs, since less transmitting nodes might be needed to reach sufficient levels of performance.

1.2. Scientific Relevance

As it will become clear in Section 3.1, so far, when the subject of study is to identify an entity within space, the leading interest of the associated research has been essentially the improvement of the accuracy of the positioning estimations in terms of numerical coordinates, or in other words, the minimization of the difference between the coordinates of the estimated position and the actual position. Although this notion is highly applicable to various scientific fields (e.g. Surveying, Radio Navigation, etc.), there are still cases, such as the field of Geomatics, where the symbolic enrichment of a position may be the most interesting feature (Kolodziej and Hjelm, 2006).

Even for indoor positioning systems offering highly accurate positioning coverage (e.g. sub-meter), it is easy to depict the value of grouping different points in space, into distinct spatial sets (or locations) of specific semantic properties. For example, a university student searching for the "Lecture Hall B3", would prefer making a lookup based on the room's name in a hypothetically provided indoor positioning system App, instead of some specific coordinates. In a similar way, other users having mobility impairments would recognize the worth of an indoor positioning system that supports space semantics as described by L. Liu et al., 2019, to be able to search for navigation routes via zones that are accessible by them. This importance and generally the difference between localization and positioning aspects has been acknowledged even from plainly technical sources (Karl and Willig, 2005). For example, in Figure 1.1 sacrificing the general positional accuracy to increase the general localization accuracy (e.g. the building is on the left of the river and not the right) might be preferable.

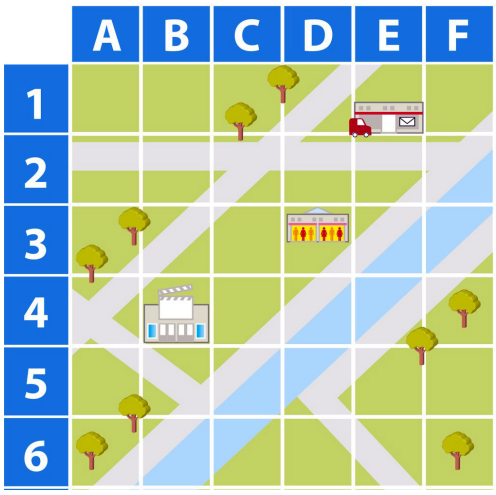


Figure 1.1: Position/Location aspects in a spatial scene
(source: Dorling Kindersley Limited)

1.3. Optimization's Test Case

Although the broader lines of our proposed optimization have already been introduced (Xenakis, Meijers, et al., 2019), in this thesis, the actual implementation is presented. For this purpose, a real indoor environment has been considered as the test case where the optimization of an actual node deployment is being examined. More specifically, this environment is an area of approximately $900 m^2$ within the Faculty of Architecture and the Built Environment at TU Delft and since our methodology involves the modeling of this environment, different model sources including the digital blueprints of the building and a recently acquired point cloud (Staats, 2017) have been utilized to produce an, as accurate as possible, representation of the indoor space.

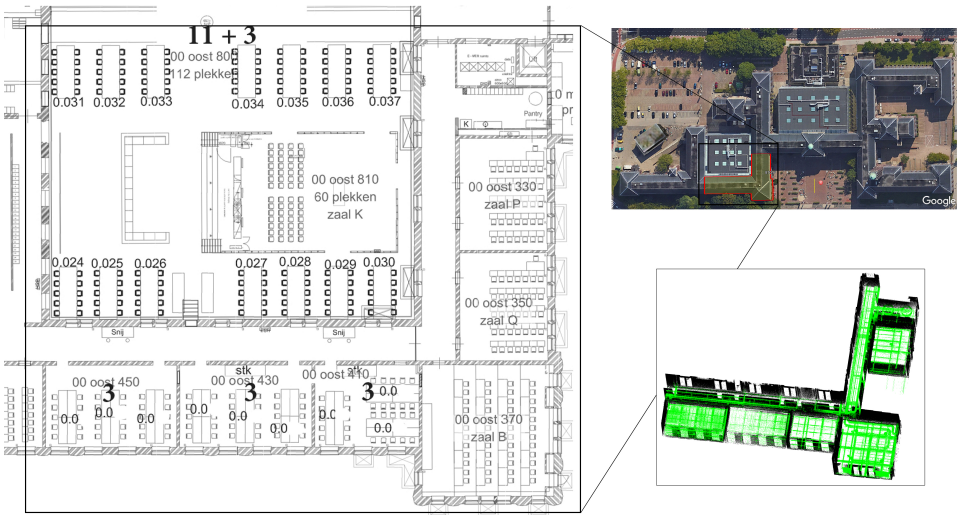


Figure 1.2: Fusion between Point Cloud model & Floor Plan
(sources: BK, Google Earth, Bart Staats)

1.4. Research Questions and Scope

As observed, literature's focus has so far been the improvement of the physical (in terms of numerical coordinates) positioning estimation. However, since positioning is only the basic part of a true location-aware solution, the objective of this research project is to cover this void by providing a direct optimization for localization purposes. Achieving that would arouse a cross-disciplinary contribution since Location-enabled Indoor Positioning Systems are not solely relevant to the field of Geomatics. Having this objective in mind, the formulation of the following main research question arises:

To what extent can the placement of BLE nodes used for fingerprint-based positioning, be optimized to increase the location distinctiveness in an indoor environment?

Increasing the location distinctiveness means, in other words, to improve the accuracy of the zone prediction among different zone-areas. However, to eventually achieve that, the following emerging sub-questions need to be answered first:

- *How can the location distinctiveness be defined for an indoor positioning system?*
- *Which metric would be most suitable for measuring the location distinctiveness among different zone areas?*
- *Which radio propagation model would offer good accuracy-complexity ratio?*
- *Which optimization algorithm should be utilized to support even large scale optimizations?*
- *How can the optimization results be evaluated?*

The exploration of all aforementioned research questions shall be done throughout the rest of this graduation project. However, this exploration is not limitless and thus, its scope should be clarified.

To begin with, as it will become clear, the major factor affecting the pragmatical optimization's magnitude is the accuracy of the simulation of the radio signals. Regarding the latter, however, several adoptions which are going to be made, introduce the corresponding limitations. For example, for the radio transmission modeling, we only consider a specific type of BLE nodes as shown in Figure 4.22 (having specific software and hardware configurations). Therefore, our modeling might be less representative for other types of BLE nodes.

Moreover, since positioning nodes are often placed on vertical walls (especially when the ceiling is high), our optimization shall consider as candidates only these positions.

Outside the scope of this project is to also find the minimum number of nodes required for a given performance. For that, it is required that we first accurately model the radio noise in the system and as it will be explained in Section 2.2, this task is highly complicated.

Finally, although this would be quite illustrative, the evaluation of our optimization shall only be performed in a simulated environment instead of a real one. Nevertheless, this does not diminish its value since evaluating the optimization in a real scenario would essentially evaluate the accuracy of the environment and radio propagation modeling, instead of the value of our proposed metric.

1.5. Thesis Outline

The rest of this thesis is structured as follows:

Chapter 2 provides the essential theoretical background that is required for understanding the subject without hassle. Different positioning techniques are briefly summarised, with the fingerprint approach being discussed in more detail. Moreover, the necessary steps composing such an optimization process are explained and justified.

Chapter 3 presents the relevant research that has been done heretofore and elaborates on different options that were available for adopting in our methodology.

The applied methodology had to respect several assumptions and constraints regarding the definition of the localization performance and the modeling of both the signal propagation and the environment within which the signals propagate. These became the foundation of the methodology and are discussed in Chapter 4. Moreover, this chapter covers the integration of artificial intelligence into the optimization's process by describing how a Genetic Algorithm (GA) became the driver of this process.

The setup and process used to evaluate the optimization's success is described in Chapter 5.

Chapter 6 presents the results of the applied evaluation and provides an analysis and an interpretation of them.

Chapter 7 elaborates on the knowledge that has been extracted throughout this thesis, applying critical thinking and answering individually the formed research questions. Finally, suggestions are given regarding future work.

2

Theoretical Background

For geo-locating an entity, one can choose between different approaches that can be classified in various ways. These include the positioning accuracy that the solution could offer (ranging from μm to km), the nature of the signals that are used to deduct the position (e.g. magnetic, electro-magnetic, sound, gravitational, etc), or even the total implementation & usage costs. BLE signals belong to the family of EM radiation at the ISM Band, and a positioning system based on them can be quite economical for the level of positioning detail that it may offer.

2.1. Indoor Positioning Techniques using BLE Signals

Depending on the nature of the signals, there are also different ways to analyse them for producing positioning estimations. Each technique may consider different signal properties (e.g. the wave's amplitude, phase, angle of incidence, travel time, etc.) introducing at the same time some corresponding constraints. For example, to be able to measure at an effective level, the travel time of a radio signal that propagates at the speed of light, specialized (and quite expensive) equipment are required. This section presents the indoor positioning techniques that can be used with BLE signals (individually or even combined), specifying also in each case, the signal's property that the method is based on.

2.1.1. Proximity Detection

Bluetooth and consequently, BLE technology has seen great success due to the underlying low hardware costs and its high suitability for numerous applications that require reliable connectivity with low power consumption. Yet, for indoor positioning and localization purposes, BLE suffers from some important drawbacks which are discussed in Section 2.2, making its effective utilization quite challenging. Because of that, better than a few-meters accuracy may not be easily achieved

in an IPS that is based on BLE signals, unless more equipment and maintenance effort is utilized. Since that translates to higher implementation costs, the most usual and also simplistic method for BLE positioning is the proximity detection.

Proximity detection works by deploying several transmitting BLE nodes at known positions and then measuring with a receiver (at an unknown position), the amplitude (i.e. the amount of energy carried) of all incoming signals. Finding the node that sent the strongest received signal, is enough for deciding that the receiver is closest to that node's known position. This technique requires that during the deployment, the positions or locations of the nodes are recorded; which is a fairly "inexpensive" process. Yet, such an IPS can basically offer valuable positioning estimations only when the receiver is next to a node.

2.1.2. Bluetooth Direction Finding

On 2019, after the release of the Bluetooth's Core Specification v5.1 (Woolley, 2019), a new positioning technique became officially available for IPSs which are based on BLE technology. This technique depends on a set of additional modules that incorporate specialized arrays of antennas that are able to identify the direction from which the signals were received (Angle of Arrival case) or the direction towards which the signals have been transmitted (Angle of Departure case). To achieve that, the system considers both the phase and amplitude properties of them. Eventually, a position estimation can be deduced via a series of trigonometric calculations as shown in Figure 2.1, a process known as triangulation which is traditionally used in the field of surveying. Depending on the topology and density of these modules, an IPS (Location-enabled or not) that relies on this method could potentially offer across the deployment area, positioning or location estimations of very high accuracy; but at an expense corresponding to the number of modules.

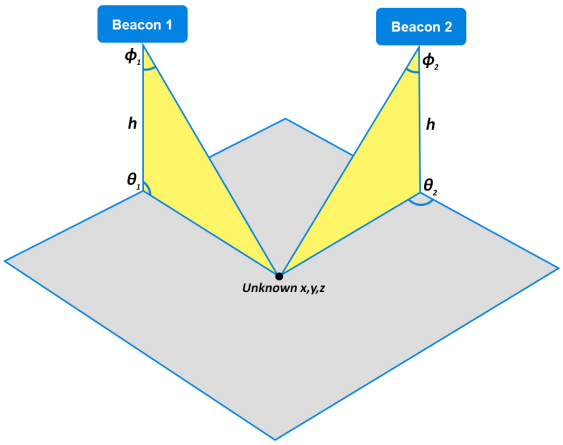


Figure 2.1: Triangulation 3D Setup

2.1.3. Lateration Using RSSI

Among the most important and generally utilized positioning techniques is the Trilateration. Its importance becomes apparent when realizing that without it, satellite positioning systems such as the Global Positioning System (GPS) could not exist, along with every other service or product that depends on it. Trilateration belongs to the greater family of Lateration techniques, the common characteristic of which is the ability to deduce positioning estimations via geometrical calculations that are based on distances and known positions, as illustrated in Figure 2.2. For calculating these distances, such systems often depend on very precise clock mechanisms that have the ability to distinguish moments in time in fine detail. Therefore, knowing that waves propagate at predictable speeds (i.e. close to the speed of light) and by measuring for how long a signal has been traveling, the system can estimate a distance.

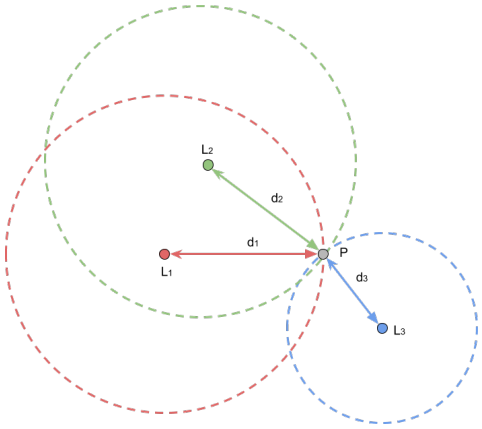


Figure 2.2: Trilateration 2D Setup (source: Alan Zucconi)

Trilateration as described above (i.e. based on timing signals) has been proved to be very successful for outdoor spaces. Yet, for indoor spaces, where there are many obstructions that can reflect a signal, this approach becomes quite problematic. Reflections lead to a problem called multipath where, as the name suggests, signals are reaching the receiver from many different paths and after various propagation times, making it difficult for the receiver to identify the true origin of the signal and thus, the real distance from it. Although there are radio technologies that are (at some level) capable of solving the problem of multipathing in an indoor environment, such as the Ultra-Wide Band (UWB) standard (Kolodziej and Hjelm, 2006), BLE still suffers from it. For that reason, this type of signals cannot be effectively timed and so, Trilateration requires another way of deducing these distances; namely, the attenuation to distance modeling which is a technique attempting to express the distance that a signal has traveled, as a function of its Received Signal Strength Indication (RSSI).

The trilateration technique, as suggested from our research question in Section 1.4, is not the one being considered in this project. Yet, research projects exist with an objective close to ours (i.e. positioning optimization via deployment adjustment of BLE nodes) which do consider the trilateration technique. Taking an example of them (Haagmans, 2017) it can be argued that among their strong points is their direct applicability to other (even more accurate) technologies such as the UWB.

At this point, it should be noted that, although the attenuation and distance properties are indeed highly related, in practice, numerous other factors which are difficult to model and that also affect the signal's attenuation (e.g. reflections, refractions, interferences, antenna geometry, etc.) act as noise, introducing to this model a lot of uncertainty.

2.1.4. RSSI Fingerprinting

RSSI Fingerprinting is the last, but also the most utilized technique in IPSs which are based on BLEs signals. It is a powerful classification technique that is generally applicable to a broader category of problems where pattern-matching of sensed signals is involved. As it shall be explained, it does introduce several important requirements in terms of time consumption for its implementation; yet, among its strong points are the facts that:

- it can be easily combined with different types of signals (leading to more accurate and reliable estimations),
- it does not require any additional/specialized equipment or even to be aware of where the signals are coming from,
- it reduces the need for modeling the sources of noise and
- its effectiveness is at some degree related to the amount of time that has been dedicated for its implementation.

Fingerprinting technique for indoor positioning applications considers the attenuation of the signals and is practically divided into 2 phases; the training phase and the evaluation phase. During the training phase and assuming that a set of transmitting nodes is deployed in the area, a receiver records at different sample positions a) an identifier for this position (e.g. its coordinates) which is typically appointed manually from the user, b) which signals can be received at that position and c) their corresponding RSSIs. The entire recording process produces a trained database (or a Radiomap) for the IPS where the fingerprinting is implemented. During the evaluation phase, a new reading is received at an unknown position, recording this time only the b and c parameters. Then, depending on the positioning algorithm being used (e.g. Bayesian Methods, k-nearest Neighbors (kNN), Neural Networks, etc (H. Liu et al., 2007)), the system attempts to classify the new reading based on the Radiomap.

In the proximity approach which was discussed in Section 2.1.1, although there is no training phase, the position of the nodes is known. In that case, the system can offer reliable estimations for unknown positions that are close to these nodes. In this case, however, although the placement of the nodes is not known, the sample positions that were used during the training phase, are known. As such, the system can offer reliable estimations for unknown positions that are close to the positions that were sampled from the user during that training phase. This means that, potentially, more detailed (and time consuming) training phase, could lead to better performance for the IPS.

2.2. The Challenge of RSSI-based Positioning

Self-evidently, for the RSSI fingerprinting technique to work best, it is important that the conditions affecting the radio propagation in the indoor environment remain as invariable as possible. For example, if after the training phase, a transmitting node was moved to another place, then this would introduce a lot of ambiguity to the positioning method. However, achieving perfect radio consistency between the sampling procedures is in practice not possible due to the numerous factors that affect the signal's attenuation. A list of such factors can be very long including moving obstacles and absorbers (e.g. people walking), changes in furniture, interference from other radio devices, etc. Yet, even if all these factors were eliminated from the equation, as it has been shown (Xenakis and Verbree, 2018), there are two others that are almost impossible to predict. Namely, the receiver's antenna orientation in respect to the transmitter's antenna orientation, and vice versa.

More specifically, Figure 2.3 presents 4 different smartphone receivers that each has been used 4 different times to measure the attenuation from a beacon. The experiment was done in a controlled environment where the only change between these 4 times was the smartphone's orientation in respect to the transmitting node (i.e. each smartphone image can be thought as how the node sees the smartphone). As illustrated from the four plots, even by changing the phone's orientation, the RSSI distribution changes significantly. The kernels (or spikes) within each distribution correspond to one of the three (37-38-39 of Figure 2.4) channels on which the BLE standard is designed to operate.

The same plot reveals another significant problem regarding the invariability of the radio conditions between a training and an evaluation phase. It is clear that between different devices, the received attenuation differs which means that a trained Radiomap that has been produced using a specific device, would be actually less representative when an evaluation check would be done with RSSI samples from another device; which is practically what happens in reality.

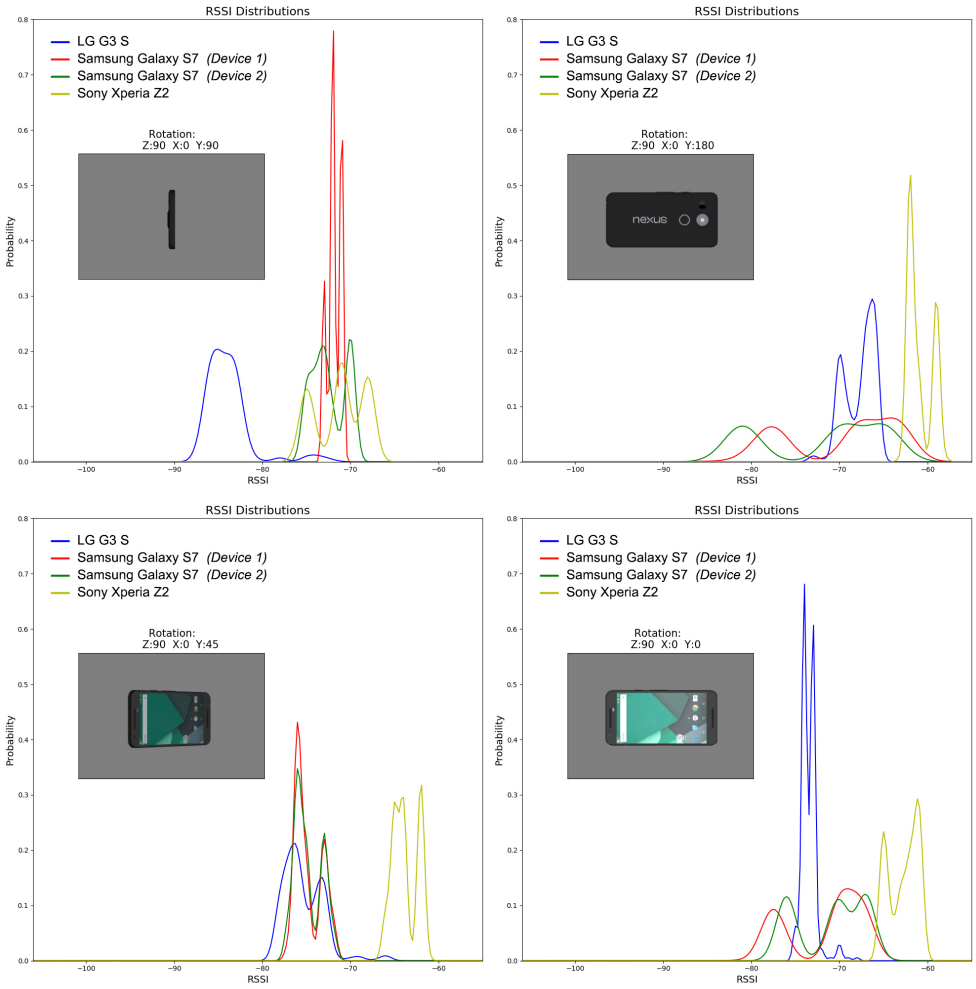


Figure 2.3: RSSI distributions of devices having different orientations

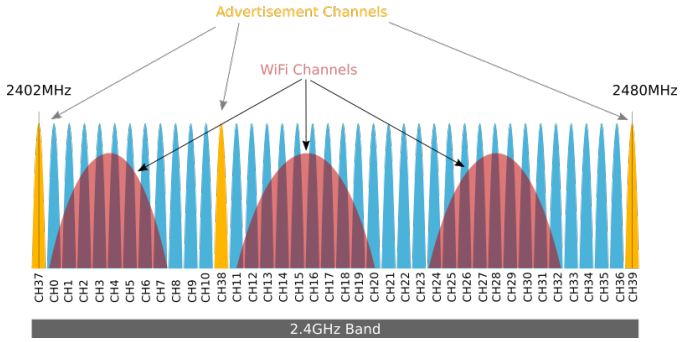


Figure 2.4: The 3 BLE Advertisement Channels in 2.4GHz Band
(source: Argenox Technologies LLC)

A usual attempt to reduce at some level the impact of having different signal responses under different orientations of a device, is to sample during the training phase many and evenly distributed (to avoid any bias towards specific ones) orientations. Although this leads to greater variances of RSSI values, it may allow randomness to smooth over time some of the noise.

Introducing another dimension of ambiguity, the same problem can be observed when the orientation of the transmitting node (in respect to the receiver) is the only parameter that changes. Figure 2.3 presents how the Gaussian RSSI distribution of 3 different devices changes unevenly, when the orientation of the node is changing. Nevertheless, it can be seen that for each device, this change remains the same across different sampling moments (Scan 1/Scan 2).

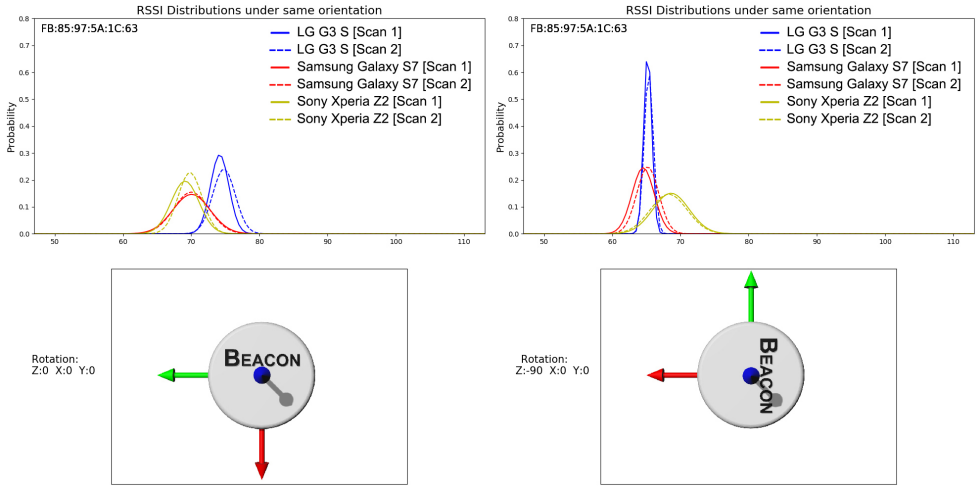


Figure 2.5: RSSI distributions of devices at 2 different node orientations

The inconsistency of the radio conditions act as noise to the positioning or localization process and is the most significant factor that affects the performance of any IPS which is based on the attenuation of the signals. However, although it is very challenging to accurately model and correct this inevitable noise, efforts can be made to reduce at some level how prone the system is to it. In essence, this is where the placement optimization of the positioning nodes is expected to play a role.

2.3. The Optimization's Workflow

To optimize the placement of a set of positioning nodes it is required that a specific workflow is respected. Without a doubt, different deployment scenarios need to be assessed and the one offering the best performance shall be selected as the best solution. However, to manually install and measure on the field all possible setups, is impossible and thus, an automated mechanism is required that will be able to simulate all these different scenarios. As expected, such a simulator needs to be aware of not only the geometry and properties of the indoor environment for which the optimization is performed, but also the physical laws affecting the radio propagation within it. Therefore, modeling the indoor environment and the radio propagation are two additional components that the optimization workflow requires.

The next step is to define a performance metric, or in other words, to "teach" the optimization mechanism how does a good localization performance look like. This metric will be applied for each different simulated scenario, leading eventually to the optimal solution.

Finally, the last step that the optimization workflow requires, is to develop a function to act as the driver for the whole optimization process. Undoubtedly, simulating in a brute-force approach all possible deployment scenarios, would certainty find the best setup. In practice, however, even with the help of the simulator, it would be impossible to assess all possible cases since these can be infinite. Therefore, the task for this function is to identify and assess the deployment scenarios which have the highest probability of offering the optimal performance.

3

Related work

Searching through the scientific literature, one can easily find hundreds of research papers broadly related to mesh optimization. In this case, however, where localization based on RSSI Fingerprinting is involved, the most relevant projects have to do with a) Wi-Fi or BLE Node placement optimization for indoor positioning purposes, or b) Wi-Fi Access Point placement optimization for any other purposes. The second category was also reviewed because BLE and Wi-Fi systems operate at the same radio band and thus, their signal propagations are very similar.

As expected, the common characteristic in each case, was the consideration of the workflow that was mentioned in Section 2.3; namely, 1) which metric of performance is being considered, 2) which signal propagation model is being used and 3) which function is implemented to optimize the chosen metric.

3.1. Performance Metrics

From all reviewed papers, most had considered (usually among other metrics too) the total coverage and signal strength in the area of interest (Adickes et al., 2002; Q. Chen et al., 2014; A. Dalla'Rosa et al., 2011; Alexandre Dalla'Rosa et al., 2008; K. Farkas et al., 2013; Fortune et al., 1995; Grubisic et al., 2009; Ji et al., 2002; Kang et al., 2013; Kondee et al., 2015; Kouhbor et al., 2006; Liang et al., 2012; Liao et al., 2011; Maksuriwong et al., 2003; Moreno et al., 2015; Nagy and L. Farkas, 2000; Politi et al., 2016; Vilović and Burum, 2014; Yoon and Kim, 2013; Yun et al., 2008; Zhang et al., 2014). That, since most applications had utilized Wi-Fi APs and so, the RSSI coverage could not be neglected even in the cases where the mesh was highly intended for indoor positioning purposes. However, the success of an indoor positioning service essentially depends on the accuracy of the positioning estimation, which does not solely depend on the total signal strength. Therefore, other metrics that are completely orientated towards indoor positioning, might be better.

On the other hand, although several research projects did use as a metric, the positioning accuracy (trying to minimize the expected error of the position's estimation) (Baala et al., 2009; Y. Chen et al., 2006; Ficco et al., 2013; Hara and Fukumura, 2008; He et al., 2011; Laitinen and Lohan, 2016; D. Li et al., 2015; Liao et al., 2011; Redondi and Amaldi, 2013; Roberto et al., 2003; Sharma et al., 2010; Voronov, 2017; Zhang et al., 2014), making an accurate prediction of the error of a positioning estimation is quite challenging, due to the complexity of properly modeling the error's sources themselves.

In fingerprint-based positioning applications, since statistical uncertainty cannot be avoided, the more discrete the estimations are, the better. According to this perception, another optimization approach is to maximize the vector distance of RSSI fingerprints in the area of interest. Although this idea has been favoured by the latest papers on this field (Alsmady and Awad, 2017; G. Chen et al., 2013; Q. Chen et al., 2014; Du and Yang, 2017; Eldeeb et al., 2018; Meng et al., 2012), it still has some points of criticism, mainly related to the way that this distance is statistically measured.

Other performance metrics that have been suggested in literature, measure the Channel interferences (Moreno et al., 2015; Wertz et al., 2004), and the Signal-to-Noise Ratio (SNR) at the area of interest (Fang and Lin, 2010; Rengarajan and Veciana, 2005; Talau et al., 2013). These, however, are more applicable for infrastructures using Wi-Fi nodes.

It is worth mentioning that in many cases, the objective of the research was to also optimize the number of nodes needed to be deployed (and thus the installation cost) (Q. Chen et al., 2014; K. Farkas et al., 2013; Ficco et al., 2013; He et al., 2011; Huszák et al., 2012; Kondee et al., 2015; Kouhbor et al., 2006; Laitinen and Lohan, 2016; D. Li et al., 2015; Liang et al., 2012; Maksuriwong et al., 2003; Mc Gibney et al., 2010; Moreno et al., 2015; Nagy and L. Farkas, 2000; Redondi and Amaldi, 2013; Rengarajan and Veciana, 2005; Talau et al., 2013).

3.2. Radio Propagation Models

As already discussed, to find the best node-setup scenario for an environment, it is important to be able to simulate the radio propagation within that. For this purpose, several options are available offering different trade-offs between computational complexity and accuracy.

The simplest radio propagation models are the empirical ones that do not consider the physical geometry of the propagation environment (e.g. one-slope model, linear attenuation model, etc. (Luo, 2013)). These models offer fast but less accurate estimations for indoor environments and thus, they have mostly been avoided in literature.

Another kind of signal propagation models are those respecting stochastic processes. These assume the existence of underlying physical phenomena, such as Multipath Fading, affecting the signal propagation in a statistical manner. The well-known Log-Distance Path Loss or Log-Normal Shadowing Model (Akl et al., 2006) falls into this category and literature review showed that, so far, most papers have been using a mix between them and various considerations of the obstacles within the propagation environment.

The last type of radio propagation models are the deterministic ones, which include a) the Ray-optical models and b) the Finite-Difference Time-Domain like ones (Remley et al., 2000). These can offer the highest level of accuracy since they consider at an analytical level, both the electromagnetic properties and the propagation environments. This accuracy, however, comes with a cost in complexity (and computational load) and so, only a few optimization projects were found having adopted them (Grubisic et al., 2009; Minkara and Shepherd, 2014; Moreno et al., 2015; Wertz et al., 2004; Yun et al., 2008).

3.3. Optimization Functions

Throughout the literature review, after having defined a) which metric to optimize and b) how the signal propagation should be simulated within the indoor environment, the next step was evidently to choose a function to handle this optimization. In general, two specific approaches have been the most favored ones. Namely the use of Genetic Algorithms and the use of Simulated Annealing Algorithms.

3.3.1. Genetic Algorithms

The goal of a GA is to translate the principles of Charles Darwin's natural selection, into a recursive procedure for solving an optimization problem; an approach that has seen wide application in various scientific fields. In the case of Geomatics, a recent example would be (Saleh and Chelouah, 2004), where this technique was used to optimize the observation capability of a GNSS-based surveying network.

As shown in Figure 3.1, this procedure is mainly the repetition of 3-steps: the selection, crossover and mutation steps. Initially, a population of individual chromosomes (or optimization solutions) is generated. Then, the strongest chromosomes (or best solutions) are selected in order to be preserved or mixed in pairs, producing the next generation of chromosomes. Some of these new chromosomes are then randomly mutated (producing again a slightly different solution) to ensure that the vast search-space is explored better. This 3-step process is then repeated, until some threshold is reached. Literature review showed that this approach was the most favoured one (Adickes et al., 2002; Alsmady and Awad, 2017; Eldeeb et al., 2018; Ficco et al., 2013; Grubisic et al., 2009; J.-H. Lee et al., 2007; Maksuriwong et al., 2003; Nagy and L. Farkas, 2000; Vilović and Burum, 2014; Yoon and Kim, 2013; Yun et al., 2008; Zhang et al., 2014).

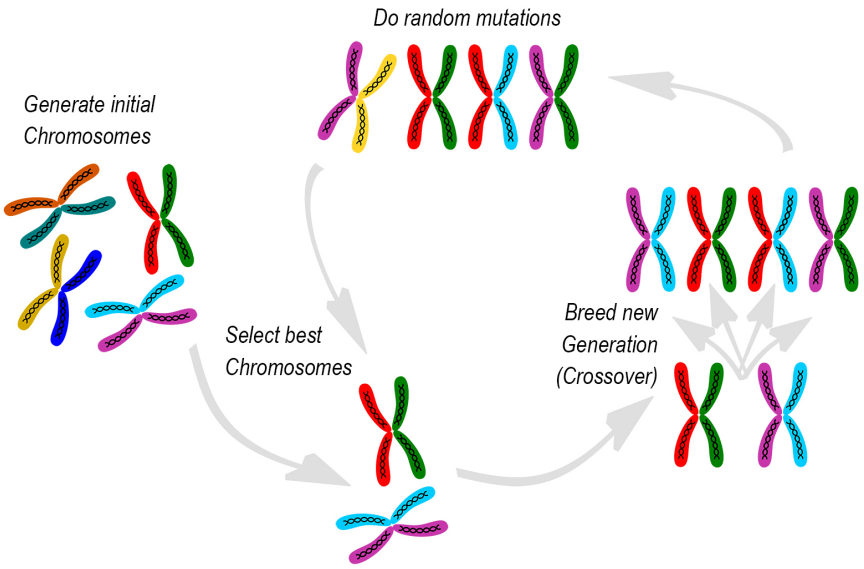


Figure 3.1: The life cycle of a Genetic Algorithm

3.3.2. Simulated Annealing

Following GA, the next most used in literature (Q. Chen et al., 2014; K. Farkas et al., 2013; Kondee et al., 2015; Roberto et al., 2003; Sharma et al., 2010) optimization method for adjusting the placement of indoor positioning nodes, is the Simulated Annealing (SA). SA is another powerful tool, that has been used across several scientific disciplines for solving various optimization problems. As an example, in the field of Geomatics and Remote Sensing, Chang et al. (2011) used this approach for optimal band selection for high-dimensional remote sensing images.

Similarly to GA, this iterative algorithm also tries to mimic an external (found in metallurgy) procedure called annealing, which helps making metals durable by heating and slowly cooling them. In practice, the searching starts with a random solution and a predefined high temperature. On each iteration, the current solution is altered and the new version is compared to the previous one, in terms of quality. If the new version is better, then it replaces the previous solution. Otherwise, based on current temperature (which gets reduced on each iteration) and the quality difference of the two solutions, the algorithm decides which one to choose. This allows for a worse solution to be chosen and thus, to escape from local optima, which means that (in the same way as GA does) the vast search-space is explored better.

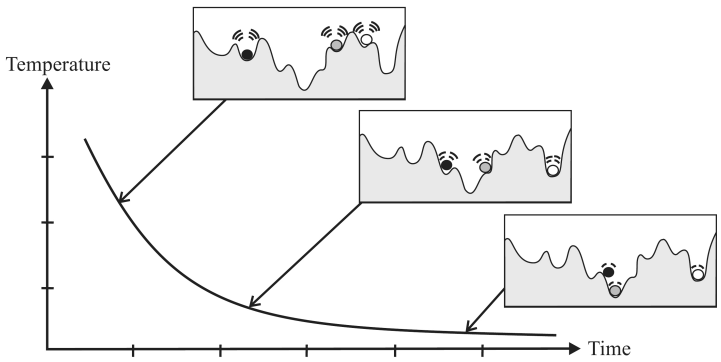


Figure 3.2: Minimization based on SA (Ledesma et al., 2012).

Figure 3.2 presents 3 solution moments of a SA optimization problem, during which, the objective is to reach the global minimum. As illustrated, higher temperatures increase the "kinetic energy" of the solutions, helping them escape local optima.

3.4. Modeling the Indoor Space

As described in Chapter 1, during this graduation project, the placement of BLE nodes used for indoor localization will be adjusted, to increase the radio distinctiveness among different zones. First, however, these zones need to be modeled. This requires a space subdivision process, which is a known problem in literature and has comprehensively been discussed (Diakité and Sisi Zlatanova, 2018; Worboys, 2011; S. Zlatanova et al., 2014). With respect to that, there are two essential aspects needed to be considered. The geometry part, since it is required for the radio simulation, and then, the semantics part, which will define the different zones.

We should, once again, clarify that our literature review has considered only research projects, the objective of which is the closest to our research question. Namely, the improvement of the localization performance via the placement optimization of BLE nodes that are used along with the fingerprinting technique. Although this thesis is the very first attempt trying to address this exact problem, yet, the same objective, but from the positioning point of view (and not the localization), has already been heavily examined. This means that optimization research efforts that considered other positioning techniques (such as the trilateration which was discussed in Section 2.1.3) were not considered since these approaches are fundamentally different and not easily comparable. With that said, out of all reviewed papers in Section 3.1, only one (A. Dalla'Rosa et al., 2011) examined during the optimization process, the 3rd dimension; and there is a good reason for that. A comparison showed that although the results between the 2D and 3D cases, were similar, the 3D case took (for a small model) 500% more time, while this percentage gets exponentially higher as the model is enlarged.

The geometry decomposition into different sections, introduces the second important aspect; the semantics. Dividing a small indoor-space (e.g. a house) into distinct zones, might sound intuitively straightforward. For example, one could distinguish a living room, a kitchen, a bathroom, etc., being separated by walls. However, what happens if walls were not there (e.g. a kitchen being connected with the living room, with no walls in between)? A problem becoming even more evident as the area increases (e.g. airports, train stations). At the same time, quite often we might be interested in separating an indoor space based on non physical criteria. For example, a museum might want to cluster different rooms into thematic zones (e.g. Paleolithic, Mesolithic, etc.).

The actual value of the optimization-product we aim to develop, can only be seen through its implementation in an indoor location-based service (LBS) which, as a final-product, would be utilizing the spatially optimized BLE nodes for localization or navigation purposes. Between these two products, the geometry and, most importantly, the semantic aspects need to be linked. For example, lets assume that the aforementioned museum was offering as a final-product, an indoor positioning service that the visitors could use to identify their locations. If this product was not aware of the aforementioned clustering, it could not take advantage of the enhancement that the optimization-product could offer (i.e to optimally distinguish the Paleolithic zone to the Mesolithic zone). On the contrary, most probably it would affect its performance. With that said, it is required that the indoor spatial model of the final-product can also support such a location awareness.

Although the development of a custom (and proprietary) model is always an option, there are already several well-established standards that could be used for modeling an indoor space. These include formats like KML (OGC, 2015), being mostly oriented towards integrations with earth browsers; Shapefile, which is a very popular GIS data format by ESRI (ESRI, 1998); GeoJSON; IFC (BuildingSmart, 2016), offering an extensive data schema for applications in the Architecture, Engineering and Construction industry domain; CityGML (OGC, 2012), designed for bigger scale modeling (cities); and also, IndoorGML (OGC, 2016). Each one of them has its strengths and weaknesses, however, among all, the IndoorGML seems to be the most powerful and suitable to be used in a final-product that could take advantage of our optimization.

IndoorGML respects several critical to our case, notions. These are the «Cellular space» which defines how the entire indoor space shall be decomposed (namely, into a set of distinct cells); the «Topological representation» which is essential for any indoor navigation application; and the «Semantic representation», «Geometric representation» and «Multi-Layered representation». The importance of these notions for an indoor model has been thoroughly discussed in (K.-J. Li et al., 2019) and without doubt, it is also directly applicable to our aforementioned needs.

4

Methodology

4.1. Defining the Location Distinctiveness

Before developing a quantitative metric that can measure how successful the localization in an IPS is, it is required that, first, a definition for this quality is given. From the main research objective, it can be inferred that the term "location distinctiveness" is used to signify this performance. Therefore, defining location distinctiveness and thus, addressing the first research question, is our initial goal.

As the term suggests, locations in an IPS can be assessed with regard to their distinctiveness. This assessment could be done using various metrics of uniqueness, however, since this research considers Location-enabled Indoor Positioning Systems where the estimations are based on radio signals, it follows that this distinctiveness needs to be assessed in terms of these signals.

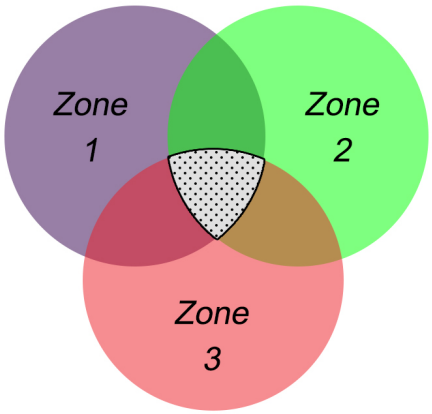


Figure 4.1: Set of positions that is a subset of more than 1 locations

As denoted in Section 1.1, the location can be considered as a superset of positions which are all attributed with the same thematic identifier; namely, the location where they belong. However, a position can simultaneously belong to more than one locations. For example, being in the elevator, at a corridor of a building section. This notion is illustrated in Figure 4.1 where the given set of positions (black dots) is a subset of all 3 locations.

Intuitively, the relations of the locations in Figure 4.1 are perfectly reasonable. Yet, allowing for this scenario contradicts the optimization's philosophy which is to make in an IPS, a location more distinct to the other ones. For that reason, the intersections of the locations need to be eliminated via location reformation. Such an example is shown in Figure 4.2, where there is no conflict with the concept of location distinctiveness. Only then, it will become possible to select which location unions to assess in terms of their distinctiveness.

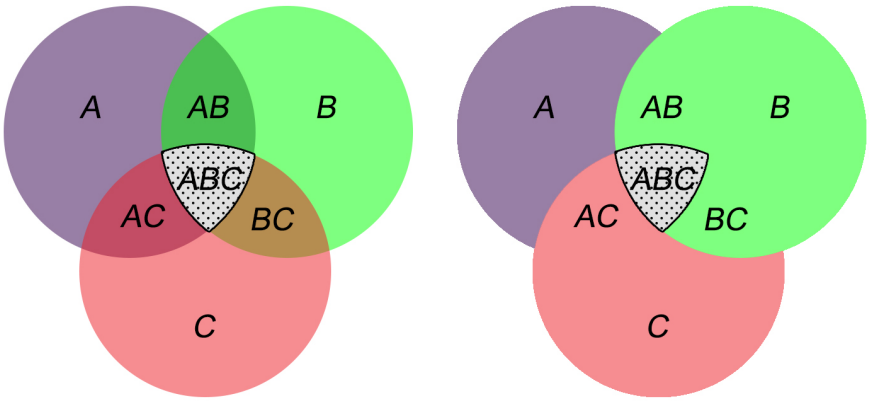


Figure 4.2: Validating the location intersections.
ABC subset belongs now to a distinct location.

Although the positions above can be described by dimensionless points, the amount of their coordinates conform with a common dimension in space to which they all (and thus, the location itself) belong. Since an IPS represents the 3-Dimensional physical space of our reality, these positions are bound to this limit. Therefore, the location distinctiveness refers to locations of at most 3 dimensions.

At the physical level, each one of the above positions is in a bidirectional binding with a unique, yet continuously varying Radio Signature (RS). Knowing some of these Position-RS bindings enables us to use positioning techniques as described in Section 2.1 and do reversed lookups based on new RSs from positions that are unknown; a lookup that eventually leads to a known position and its location. Quite often, the location where the unknown position belongs is in reality (although we are not aware of it because the position is unknown) the same as the

location where the known position belongs. As a result, the higher the probability this happens during lookups, the higher the location distinctiveness for the system is proved to be.

The factor affecting the aforementioned probability is the variation degree of these Radio Signatures. Assuming that we had two known positions, a Green and a Purple one that belonged to two different locations. Each one of these positions would be bound to a unique, yet continuously varying Radio Signature. This scenario is illustrated in Figure 4.3 where these two RS variations are presented using two probability functions. From the first plot (on the top), it is clear that the most probable RS at the Green position (the peak of the Green function) is on the left side (i.e. lower in Radio Signature units) when compared to the most probable RS at the Purple position.

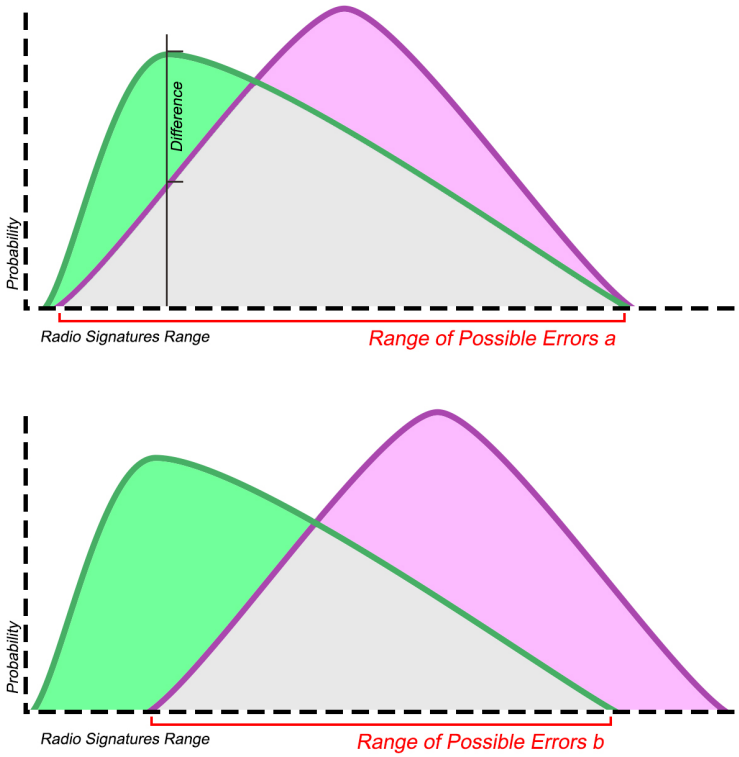


Figure 4.3: Changing the separation distance of the Radio Signatures

Let's assume that we have a new RS taken randomly from one of these two positions above (i.e. although we are not aware of it, it could either be the Green or the Purple position). Let us also assume that this RS is the same as the RS at the

Green Peak above. Using the two probability functions, we can make a reversed lookup based on this new RS. As the vertical line shows, the probability that this RS is found at the Green position is higher than the probability that this RS is found at the Purple position (with some probability difference). This could be an indication that, probably, the new RS belongs to the location where the Green position also belongs. However, this may not always be the case. Due to the variations, it is possible that, in fact, this new RS corresponds to the Purple position and thus, the location where that belongs. The RSs where this error could happen, are the ones at the intersection of these two probabilities. The bigger their range (which is indicated with the red marking in the plots), the less distinct these locations are which translates to a worse location distinctiveness.

Let's assume now that the noise which was responsible for the variations, remained the same. It would then be better if the probability functions were repositioned in a way to reduce this range of possible errors (second plot). Therefore, the overall (from all possible positions that belong to any of the locations) RS span where localization errors are not possible, defines how location distinctiveness can be perceived. That clarifies that the location distinctiveness is a performance that does not describe individually a specific location, but instead, it is used to describe, in overall, a group of locations under specific radio conditions.

4.2. Measuring the Location Distinctiveness

In the previous section (Section 4.1) a definition for the location distinctiveness was given. However, measuring this performance is far from straightforward and this complexity is due to several reasons. These shall be discussed in the following sections, leading eventually to a definition of suitable metrics for location distinctiveness.

4.2.1. Selecting a Positioning Technique

To begin with, it should be emphasized that the location distinctiveness is bound to the positioning technique which is used for the aforementioned lookups. Different approaches would produce different scores of location distinctiveness. Therefore, it is important that, first, a positioning technique is selected for the placement optimization of the nodes. From the main Research Objective (Section 1.4), it is clear that this research considers the fingerprint-based positioning technique. Yet, as mentioned in Section 2.1.4, there are various algorithms that can be used based on this technique.

In Section 2.2 the problem of modeling the noise was described. A problem that prevents us from being able to know the variations of the Radio Signatures at each position. As a consequence, algorithms based on probabilistic approaches could not easily be utilized for the optimization and thus, they shall be avoided. This means that for any single position, a non-varying RS shall be appointed. To pre-

vent this limitation rendering the utilization of the location distinctiveness unfeasible, we need first to adopt the hypothesis that by adjusting the positions of the nodes (and thus the Radio Signatures of the positions), the underlying RS variation at each position (which we cannot model) will remain unchangeable. This way, we can axiomatically say that, regardless the placement of these nodes, the further the RS distributions in Figure 4.3, the better for the location distinctiveness. If this hypothesis was not made, then it could be possible that spreading the distributions by changing the node positions, would lead to bigger ranges of possible errors, as depicted in Figure 4.4.

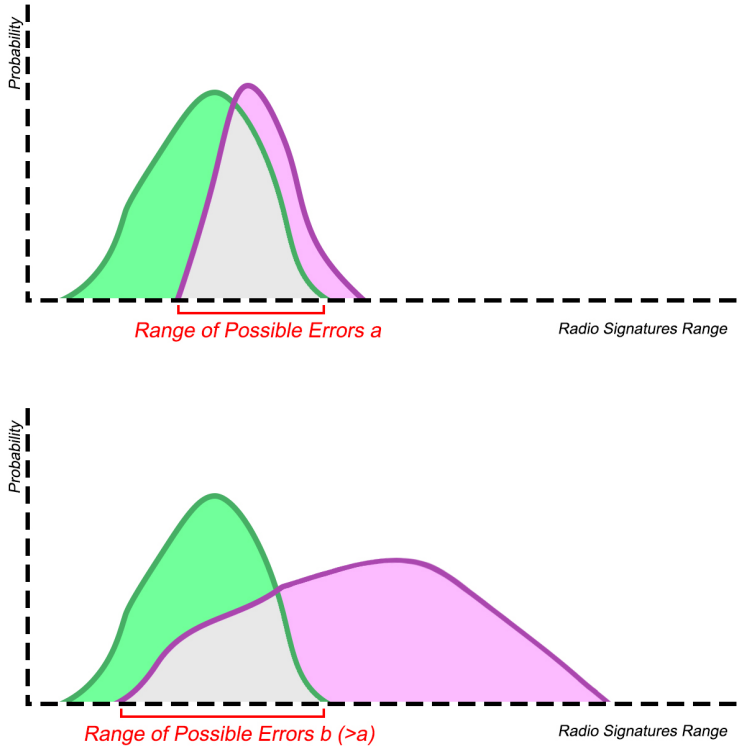


Figure 4.4: Changing the separation distance of the Radio Signatures (with changed variance)

An algorithm which is quite suitable for the purpose and is also in accordance with the related literature review in Section 3.1, is the Euclidean distance of the RSs. Therefore, this approach shall be ultimately selected and implemented as follows:

Let us denote by N_{PN} the number of BLE positioning nodes in the indoor environment and by N_{KP} the number of the known positions that belong to some location. Each one of these known positions is in a binding with a unique Radio Signature,

an RSSI vector, which is composed by the average attenuation of the signals that arrive at that known position from each of the nodes. We can denote these attenuations by $A_{(pn,kp)}$ where pn is the node from which the signals are arriving and kp is the known position. During the evaluation phase, a new Radio Signature is captured at an unknown position, containing the RSSIs from all the BLE nodes. These RSSIs are denoted by RS_{pn} where pn is the node from which the signals are arriving. The final step is to find which kp in N_{KP} offers the minimum Euclidean distance between the RS_{pn} and the $A_{(pn,kp)}$. A distance calculated as follows:

$$\sqrt{\sum_{pn=1}^{N_{PN}} \left| RS_{pn} - A_{(pn,kp)} \right|^2}$$

It should be noted that in case a node cannot reach a known or unknown position, an RSSI of -100dBm is used.

4.2.2. From the overall Location Distinctiveness to a representative one

The next challenge involved in the process of measuring the location distinctiveness is the fact that according to its definition, all possible positions should be considered. However, since space is continuous and there is no analytical way to achieve that, a representative set of positions shall be used for this measurement instead. It is important to state that by doing that, we assume that the overall location distinctiveness can be approximated by using a finite set of positions within the assessed locations. An assumption that may allow for critical thinking.

Finding these positions and implementing an algorithm which can use their Radio Signatures to compute the system's location distinctiveness, introduces the need of converting the continuous space into a discrete one. Such a conversion can be illustrated through the following group of figures (Figure 4.5, Figure 4.6), where an example of a 2-node setup within a 2-dimensional indoor environment is used.

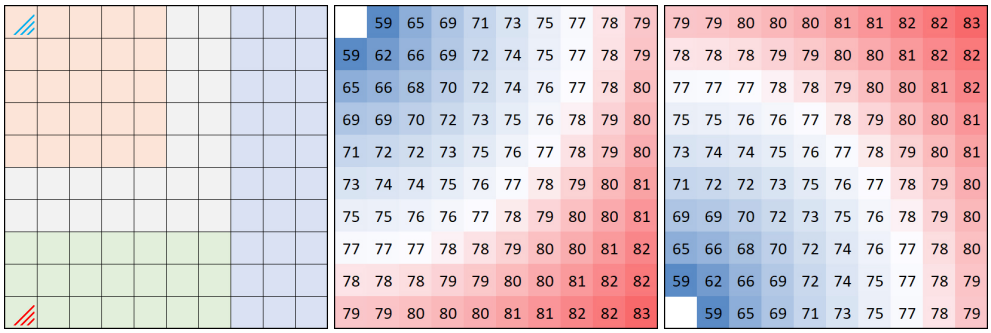


Figure 4.5: Signal coverage under unobstructed propagation

More specifically, the left part of Figure 4.5 illustrates an open (no walls) space that consists of 4 different zones (i.e. 4 locations). These are divided by a grid of sub-space cells having a total resolution of 10x10. Each cell can be defined by 1) a centroid position where a specific Radio Signature exists, b) its cell size and c) the zone where it belongs. Although the following does not apply in the continuous space of reality, in current discrete space, the aforementioned RS is allocated to the entire cell. At the corners, 2 transmitting nodes (blue & red) have been installed and their radio coverages have been simulated (based on a simplistic radio propagation model) and presented on the right parts of the figure. Since no walls exist, the signal propagates unobstructed and thus, very smoothly.

Figure 4.6 presents the same example, but this time, including walls (obstructed scenario); a difference being used to emphasize the impact of different obstruction scenarios to the radiomap.

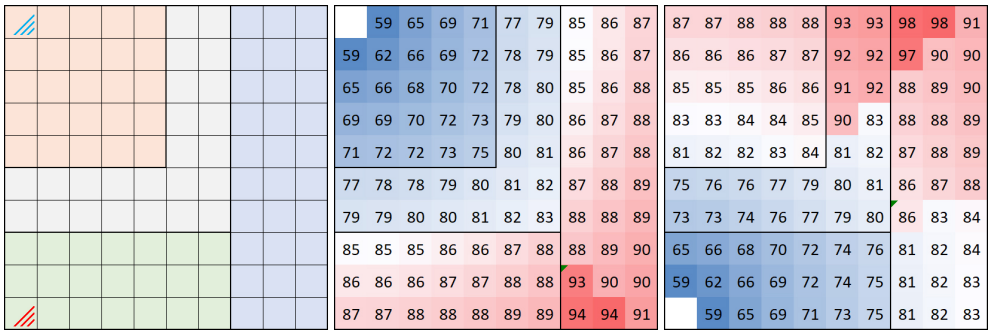


Figure 4.6: Signal coverage under obstructed propagation

At each sub-space cell, the combination of the 2 signals produces a distinct vector of 2 RSSI values. Namely, the Radio Signature at that position. Plotting the vectors of the obstructed scenario (Figure 4.6) would result in Figure 4.7, where each dimension corresponds to a specific node. Therefore, illustrating a 3-node setup would result in a 3-dimensional graph, while a bigger setup would require a hyper-dimensional representation. Grouping the RSSI vectors by their zones can help us approximate (in the RSSI vector space) the overall Radio Signatures which according to our hypothesis in Section 4.2.1 (i.e. the further the RSs, the better for the location distinctiveness), we need to spread. In the figures below, these can be found colored respectively using an approximated alpha shape. Every position within a shape belongs to the corresponding class (i.e. zone), while, all positions that are outside these shapes (separation space) belong to no class.

Figure 4.7 makes it clear that the centroid positions which define the physical border of the zones, also define the borders of their Radio Signatures. Therefore, these can become the representative set of positions which shall be used to measure the location distinctiveness. Ultimately, this association between the physical and vector space enables the transmutation of increasing the location distinctiveness, into the task of increasing the (n-Dimensional) Euclidean space that lies between them.

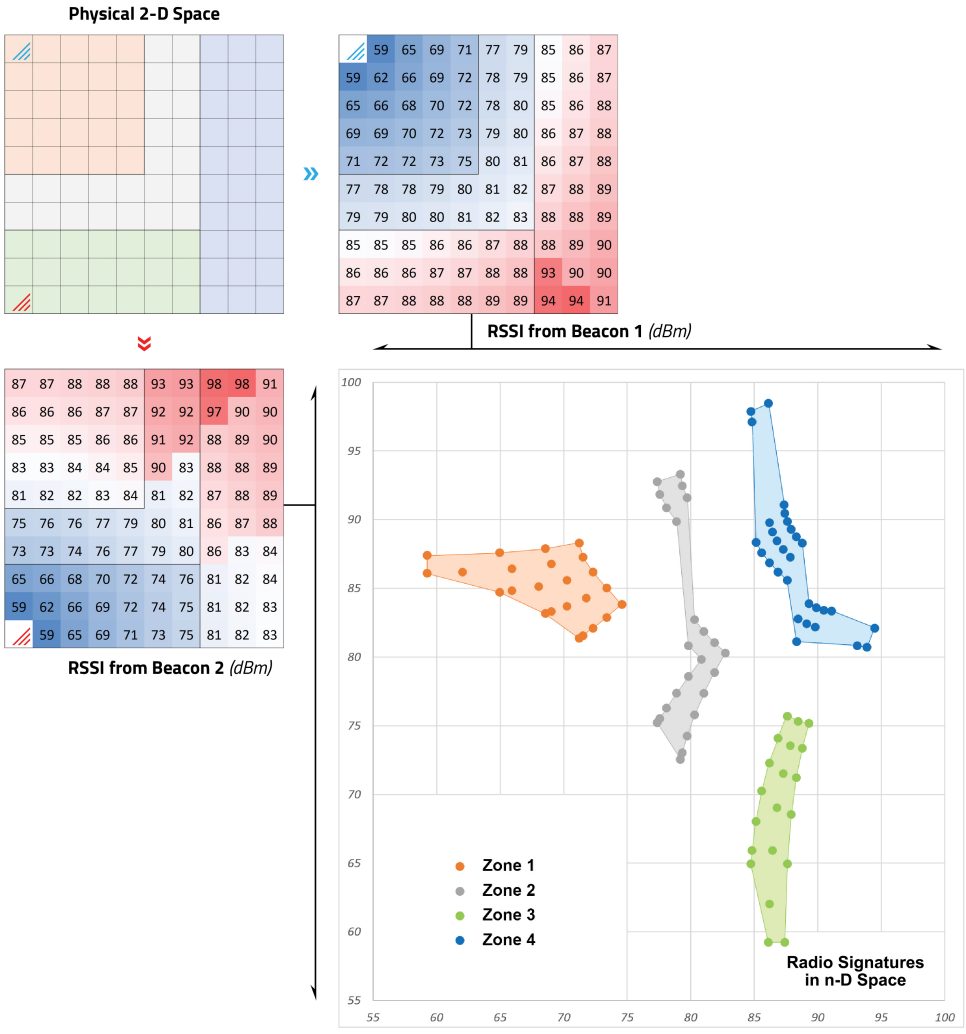


Figure 4.7: From the Physical 2D Space to the Radio Signatures

4.2.3. Modeling the Separation Area

To increase the Euclidean area between the classes of the Radio Signatures (in RSSI space) which are shown in Figure 4.7, may seem like a fairly straightforward task. Yet, this introduces an additional level of complexity regarding the way this area (its shape) is defined. As already discussed, the topology of the zone borders between the physical and the RSSI space are the same. Therefore, the borders of the area that we are after shall be computed based on the physical shape of the zones.

To begin with, computational geometry offers various different methods on how the borders of this physical area could be defined, either in two or three dimensions. One approach could be to first calculate the Convex Hull containing all the physical zones which, at the next step, would be excluded from it, leaving only the separation area. Although this approach would be quite sufficient for cases where the indoor environment is coherent (like the indoor environment of our example in Figure 4.6), yet it would not be effective for physical spaces that have irregular shapes, such as the one presented in Figure 4.8. The problem is that the physical separation area that the Convex Hull would produce in this case (grey area), would result in a radio separation area that is much bigger than the effective radio area that the positioning utilizes; and that, would lead to an optimization process that is highly "wasted" in separating Radio Signatures that are already effectively separated (e.g. due to the big distances outside the building).

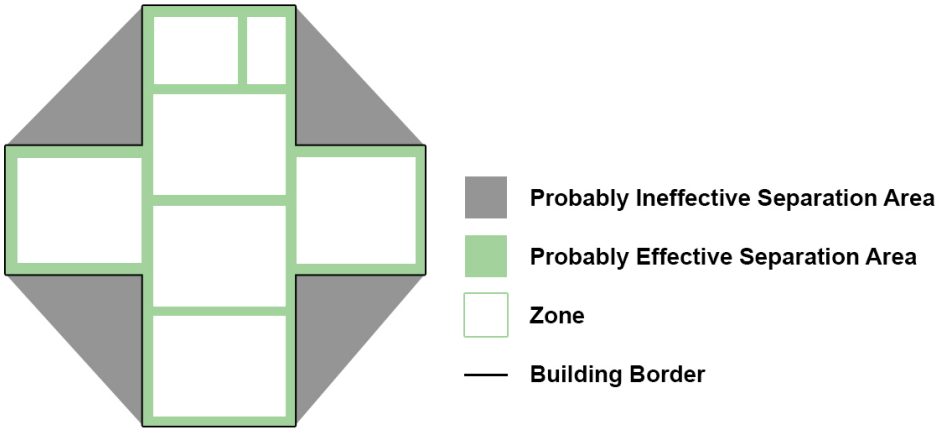


Figure 4.8: Physical Zones having irregular Shape

The solution is to calculate an alpha shape within which, separating the Radio Signature is meaningful. In other words, a shape that results in an effective separation area. With that in mind, a methodology was developed that respects the Manhattan topology of the cells as depicted in Figure 4.9. According to this method, the Alpha-Shaper starts from the left-most cell and procedurally encloses the cells

(ensuring that no cell is left outside the Alpha-Shape) by prioritizing first the "a" directions and secondly the "b" directions. Eventually, after having parsed all 4 directions, a perimeter such as the one shown in Figure 4.10 has been produced.

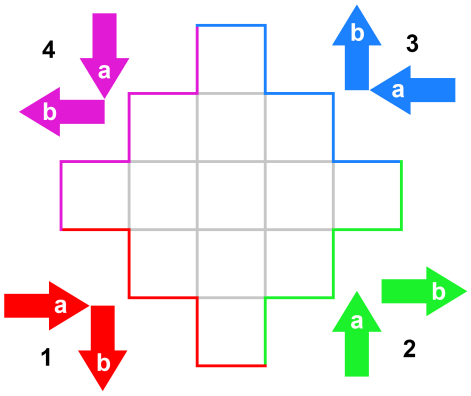


Figure 4.9: Circular Manhattan-Alpha-Shaping Procedure

From there, the next step for calculating the physical separation area of the zones (i.e. locations) is to identify the cells that are part of each zone's border. These cells are the ones which are not connected with another cell that belongs to the same zone, at all their cardinal directions (top, bottom, left, right). In Figure 4.10, these cells are depicted using stronger colors.

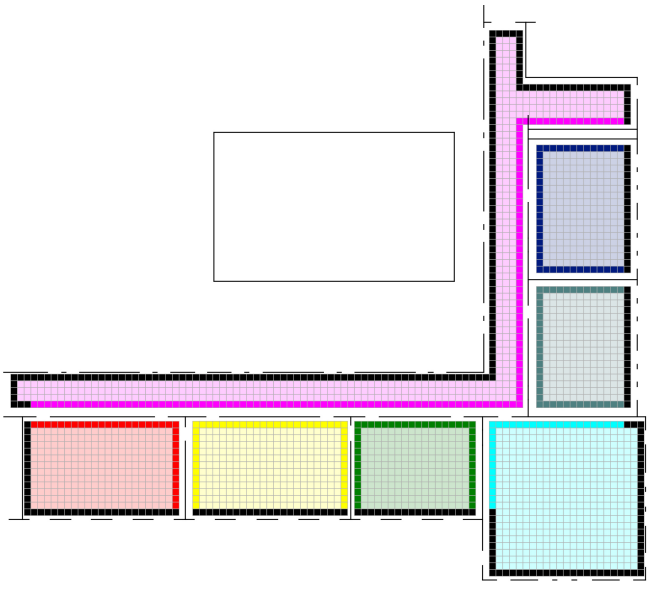


Figure 4.10: Example of a Zone Perimeter (black cells) produced by the Circular Manhattan-Alpha-Shaper

Having generated the above features and via their differences, we can compute the physical separation area. Such an example is presented in Figure 4.11 (dark area). Therefore, the topology of this area can now be used to define the radio separation area (in vector space), which the optimization process aims to maximize.

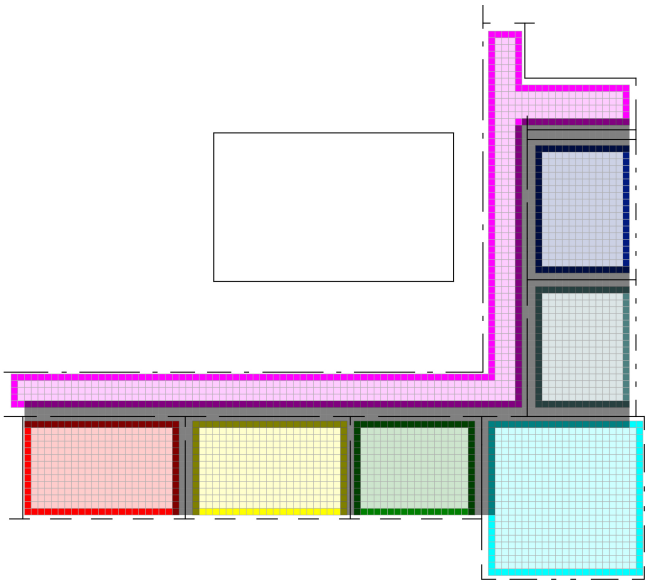


Figure 4.11: The final Separation Area

4.2.4. Modeling the Separation Distances

Up to this point, our methodology has led to the generation of a radio separation area (i.e. a hyper-plane in the n-Dimensional vector space of Radio Signatures), the topology of which is defined by the borders of the locations (i.e. zones) in the 2-Dimensional physical space. Our optimization's objective is to maximize it. However, assuming that such an area may have 100 dimensions (due to the use of 100 positioning nodes), the integration required for the computation alone of such surface, before even starting to iteratively repeat this calculation a few million times until we converge to an optimal solution, suggests how preventive it is. Therefore, another indirect approach is required, introducing, of course, the corresponding margin of inaccuracy.

Probably, among the most rational ideas would be to perform a dense triangulation (e.g. based on Delaunay triangulation) using as points, the cells. This would reduce the problem into calculating the sums of every produced 2-dimensional triangle which lie in the n-Dimensional space. However, after benchmark experiments, the computational cost of this calculation was proved to be still very demanding for the optimization's purposes. Therefore, another more simplistic idea was approached.

An area of a function is related to the accumulation of a set of specific distances within it. This notion is close to Riemann's sum where these distances gain some width and can be used to approximate the integral of this function via a finite sum (Nykamp, 2019). Taking advantage of this deduction, we introduce the concept of the "Separation Distances" for those cells that are on the border of a zone and within the separation area, to act as the feature being considered during the optimization. For each one of these cells which belongs to some zone, this distance can be defined by the connection with the physically closest to it cell that belongs to another zone. Applying this rule to the above used examples, result in the separation distances that are presented in Figure 4.12 and Figure 4.13 respectively.

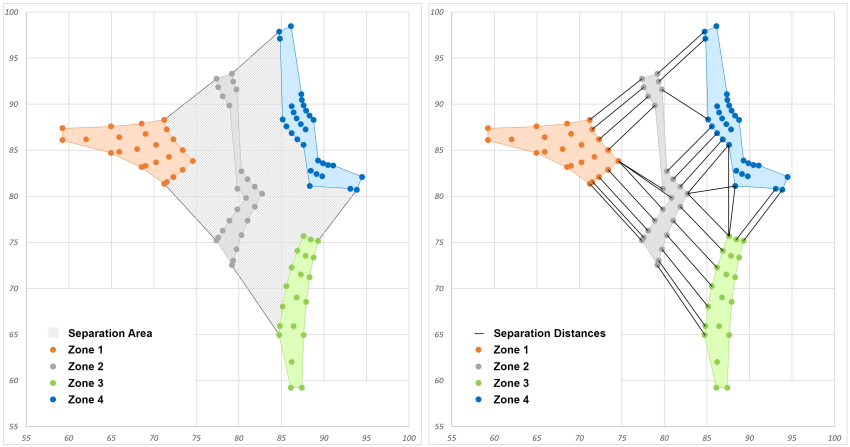


Figure 4.12: Separation Area & Distances (Example 1)

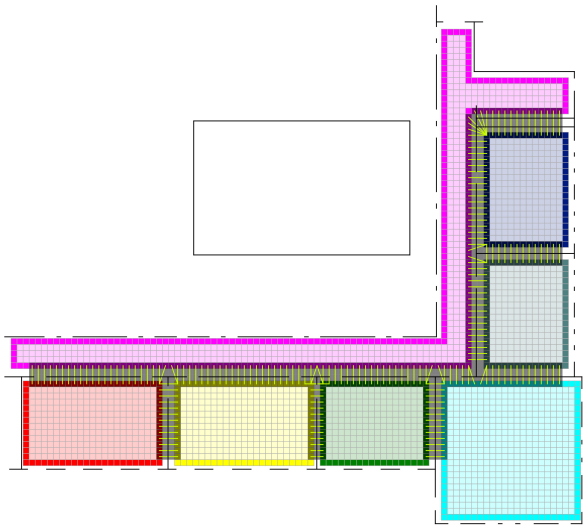


Figure 4.13: Separation Distances (Example 2)

The strong points of this approach are obvious. Firstly, to calculate an Euclidean distance is computational as inexpensive as it can gets. Also, decomposing the area to sub parts allows us to locally identify where the separation is at its highest or lowest. Moreover, this approach is implementable in every case, regardless the cell size (Figure 4.14) or the complexity of the indoor environment (Figure 4.15).

However, the introduction of these separation distances has also two important down sides. To begin with, measuring the length of an interconnection which is practically a fixed-dimensional feature, misses capturing the entire and true length of the real surface. It would be like measuring the shortest distance between two points on a 3-dimensional hill, but instead of respecting its 3D curvature, the measured distance would have considered the direct path (i.e. travelling through the hill). Yet, the closer these distances physically are (which depends on how we define/draw them), the smaller this curvature is, leading to less errors. The second disadvantage of this approach is that it introduces a bias that is related to the physical corners of the zones. At these positions, the distances become more densified, which can be noticed from the examples below.

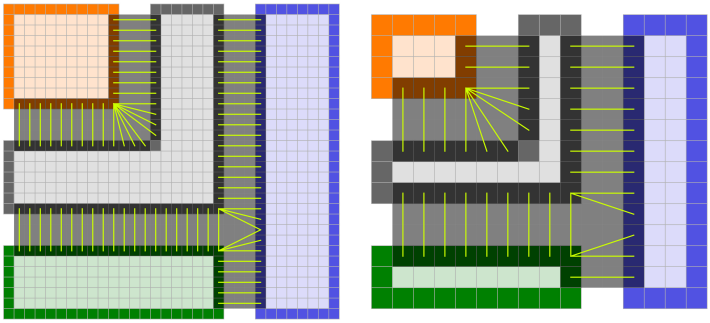


Figure 4.14: Class Interconnections at different Cell Sizes

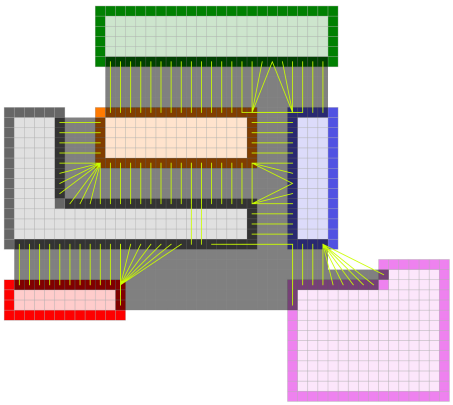


Figure 4.15: Class Interconnections at a more complex Scenario

4.2.5. Maximizing the Minimum Separation Distance

As emphasized in the previous chapter, one of the strongest points of the "Separation Distances" is the fact that it becomes possible to identify where the separation is at its highest or lowest. Therefore, we can assess different deployment scenarios and ultimately opt for the one, within which, the minimum separation distance was found to be the biggest across all of the scenarios. The pair of cells where this connection is situated, have the highest probability of offering wrong estimations in the Indoor Position System. Therefore, by opting for any other solution would lead to an even worse localization performance at that region.

In accordance with the above, to maximize the location distinctiveness using this metric, the optimization needs for every different deployment scenario to measure all distances and extract the minimum one. In the end, among all extracted minimum distances, the biggest one would be the optimal solution in overall.

4.2.6. Maximizing the Product of the n Shortest Separation Distances

The metric we defined in Section 4.2.5 ensures a minimum performance at the region where the localization is the most problematic. However, it does not guarantee that this deployment also offers the highest Radio Signature separation across the entire environment. This can become more clear using the examples presented in Figure 4.16, where 3 separation distances (red, green, blue) are shown for 3 different deployment scenarios.

In this example, we can make the assumption that the only separation distances within the indoor environment are the 3 colored lines (rgb). Therefore, maximizing the overall localization performance means that these distances need to become as long as possible. The distances corresponding to the optimal deployment that was found using the above metric (i.e. the maximization of the minimum separation distance or MMD), is represented by the first case. There, the red separation, which is the lowest in that optimal deployment, is still bigger compared to the other deployments. Yet, although in the next deployment scenario (which has no name), the red separation is shorter compared to MMD, the other two distances are remarkably higher. As a result, for an overall performance increase, all (or at least most) of the distances need to "somehow" increase as much as possible.

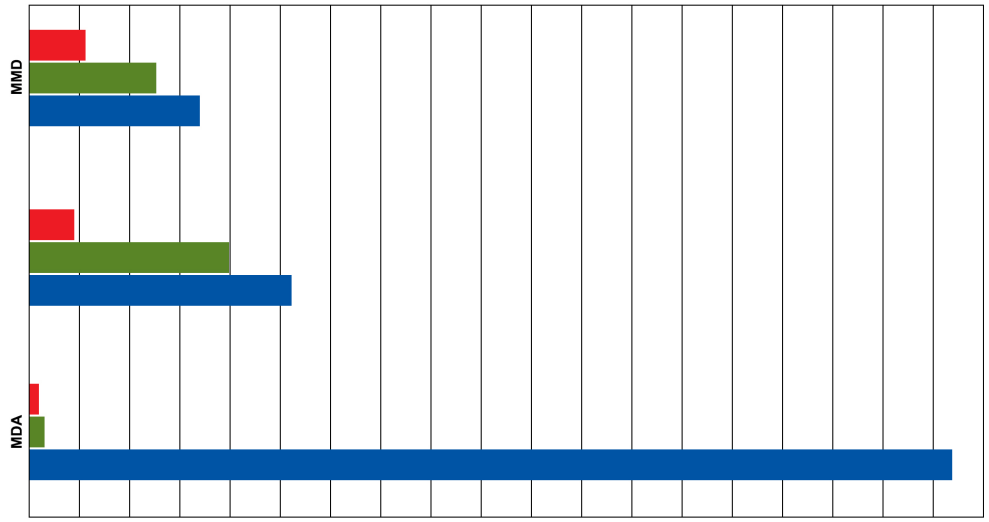


Figure 4.16: Separation Scenarios

The problem here, however, is that the increase of each distance is reversely and in a non linear way dependent to the increase of the others. Given this, a naive approach would be to compare the sum of these distances across different scenarios and, eventually, opt for the maximum one (i.e. Maximum Distance Accumulation or MDA). From the figure, it becomes clear that this metric would lead to a deployment where although the accumulated distance is maxed, the performance gain has been transferred only in a single region (i.e. the blue one). Intuitively, it would be like placing all nodes next to each other in one room. Therefore, to increase the overall distances as much as possible and in an way that the shortest distances are not diminished, is among the biggest challenges for our optimization.

With that in mind, along with the Minimum Separation Distance, an alternative performance metric is also examined. Namely, the Product of the n Shortest Separation Distances. Depending on the parameter n, the maximization of these two metrics can lead to an optimization that is quite similar (or identical in case n is 1). However, the larger the n, the more weight is given to the less problematic distances, leading probably to a separation that is in overall bigger.

The reason why the product has been considered here (and not the summation for example) is because it has the property of favouring big values and disfavouring values that are approaching zero; which is very close to the reasoning behind this metric. Moreover, as expected, the parameter n is the most important factor for this metric and its value should better be defined after tryouts. Yet, an educated guess for that value would be to reflect the amount of interconnections that are expected to be the most problematic ones.

4.3. Modeling effectively the Indoor Environment

The indoor environment of the examples that were used in Section 4.2.2 (regarding the conversion of the continuous physical space into a discrete one), was two dimensional. Yet, as mentioned in Section 4.1, the location distinctiveness and thus, its optimization, can refer to physical locations of up to 3 dimensions. Such an optimization mechanism has to begin with properly modeling the indoor environment, along with the zones of interest, as discussed in Section 3.4.

Starting with the first aspect, a major decision needs to be taken regarding the dimensions of the model. To maximize the radio distinctiveness among different zones by adjusting the node placement, one needs to be able to simulate the radio propagation within the indoor environment. Therefore, since an accurate radio propagation model requires the utilization of an accurate representation of the propagation space, the more detailed this indoor model is, the better. In theory, a point cloud-based 3D model that would include even furniture surfaces, would perform the best. However, since the model's complexity affects highly the speed of the optimization, a more efficient approach is needed.

As mentioned in Section 3.4, in a similar research project where both 2D and 3D indoor models where examined, the 2D approach outperformed the 3D one. Therefore, for the optimization, a 2D indoor model shall be used. Although opting for this approach means that we practically neglect the exact geometry of any windows, doors, or half wall openings, leading to the corresponding margin of errors, yet during the development of the radio propagation model that shall be discussed in Section 4.4, the nodes and receivers where all placed at the same height.

The need to decompose the geometry of the 2D model defined above, into different zones, introduces the second aspect; the semantics. However, considering that the zoning process may not always be straightforward, it needs to be defined.

With respect to the aforementioned, we apply the following 4 rules:

- **Assigned cells must be completely enclosed by their zones:** A cell shall be assigned to a zone only if it is, physically, entirely contained within that zone. This means that the grid's cell size highly affects the way zones are considered during the optimization.
- **Zones must not overlap:** A physical position or area should not belong to different zones, since that would contradict with the zone distinctiveness notion.
- **The interior of each zone must be continuous:** Having zones that are discontinuous introduces impracticality to their utilization and thus, it should be avoided.
- **Zone's borders must be perpendicular to the reference grid:** During the optimization process, the indoor model (i.e. the obstructions along with the zones) need to be spatially indexed into a reference grid. Ensuring that the bor-

ders of the zones are perpendicular (or parallel) to the axes of this grid (Figure 4.17) is crucial to the optimization's speed, due to the reduction of the geometrical calculations needed to be done during the radio propagation modeling.

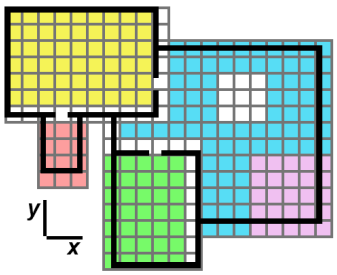


Figure 4.17: Perpendicularity of the zone's borders

It is worth mentioning that these rules allow for scenarios where the zoning is not watertight (white cells in Figure 4.17), or zones that do not follow the physical obstructions of the indoor model (e.g. North & West cells of pink zone in same figure).

Respecting the aforementioned methodology, the indoor environment of the Faculty of Architecture and the Built Environment at TU Delft (presented in Section 1.3) was modeled. Starting with the digitization of the obstructions, 3 different types were considered which can be seen in Figure 4.18. These include thick concrete walls (dark brown), thin concrete walls (red) and wooden walls (light brown).

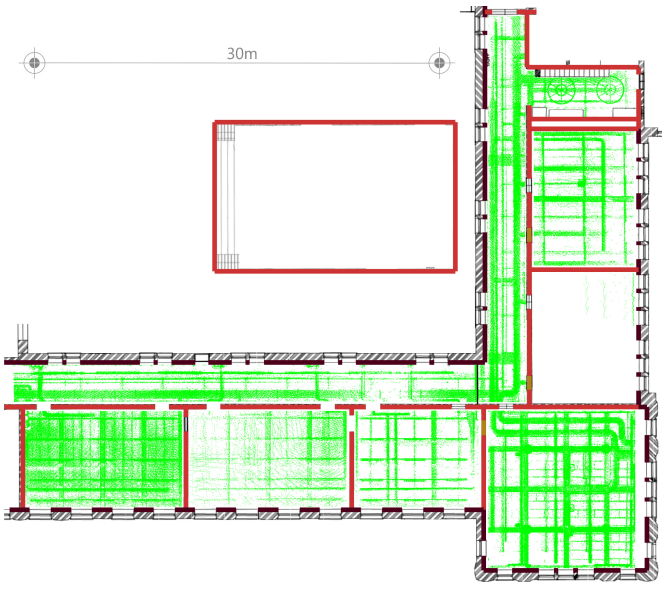


Figure 4.18: Modeled Obstructions in the Indoor Environment

The next step was to model the indoor zones (or locations), the distinction of which we are interested in maximizing. In total, 7 different zones were modeled, each representing a single room or corridor within the environment. These zones are presented in Figure 4.19 using different colors.

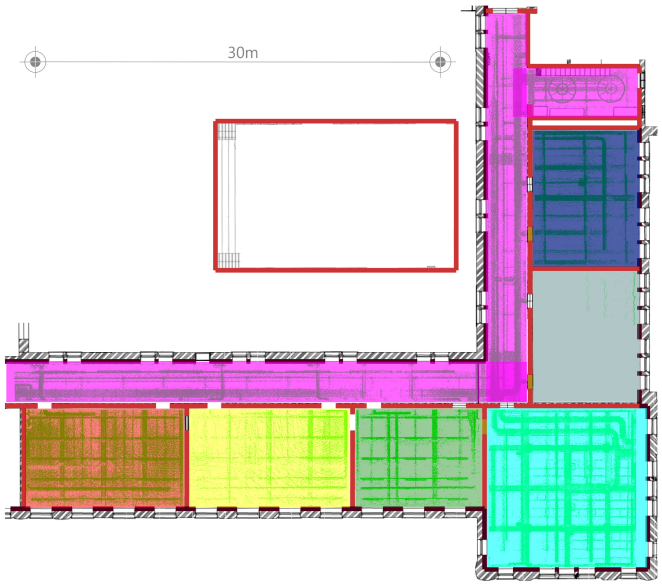


Figure 4.19: Modeled Zones in the Indoor Environment

4.4. Developing the Simulation Engine

One of the most crucial parts of the optimization mechanism is the radio propagation modeling and that, because its accuracy is directly affecting the optimization's quality and performance. As mentioned in Section 3.2, the best results require the implementation of a deterministic model and so, this will be our approach too.

More specifically, for estimating the associated power fields at every sampled cell, the ray launching technique is going to be utilized based on the Geometrical Optics phenomena of reflection and refraction (as illustrated in Figure 4.20). Assuming that each node broadcasts omnidirectional, a sufficient number of rays will be evenly (in terms of angle) generated and traced. The term sufficient is used to denote the importance of delivering ultimately (even after many reflections and refractions) the generated signal to every cell, that in reality would indeed receive it. The attenuation of each ray will be the result of a) the distance path-loss during its propagation in free space, b) the attenuation due to reflections and c) the attenuation due to refractions.

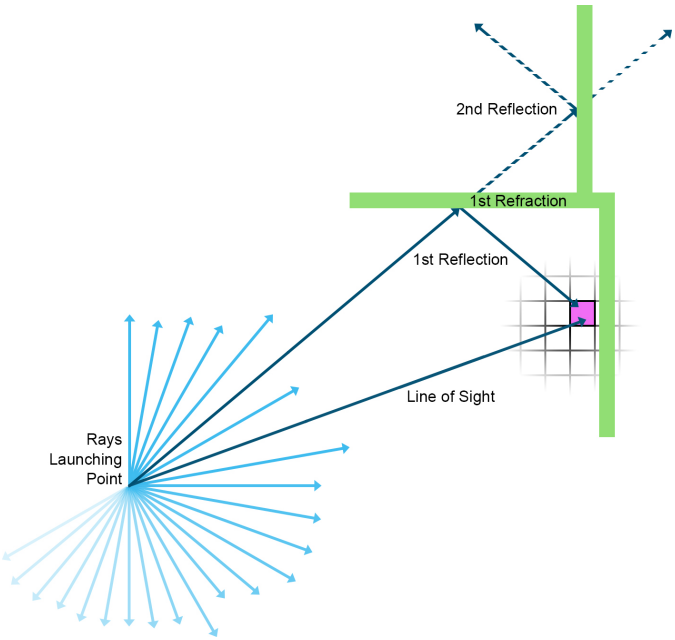


Figure 4.20: Ray Launching Example

During the simulation, several reflection and refraction coefficients need to be considered for the different types of obstructions. Although generic estimations can be found in literature, a more accurate but also much more involved approach would be to compute the optimal ones for the specific propagation environment (Figure 4.21).

Figure 4.21: Corridor View within the Indoor Environment

To find the underlying coefficients, the first step was to deploy across the area (at regular and known positions-Np) 30 BLE nodes at a height of ≈ 1.7m. Then using a 1-axis 360° rotatable mount which was also set at a height of ≈ 1.7m, we sampled 60 times (again at regular and known positions-Sp) the attenuation of every received signal. Each scan lasted 6 minutes with the receiver performing a complete rotation every 18 seconds (i.e. 20 turns in 6 minutes). The BLE beacons along with the tripod mount are shown in Figure 4.22, whereas each node and sample position can be found mapped within Figure 5.5.

Figure 4.22: BLE Beacon Set & Mount Type used for the Sampling

For every position combination Np-Sp, a set of RSSIs was collected. Then, for each one of these sets, a probability function was produced, the average of which became the representative attenuation for that combination. Since there are 30 Np's and 60 Sp's, the total combinations (or connections) are 1800. By appending these positions to the obstructions model (Figure 4.18), it became possible to identify which connections are unobstructed (i.e. there is no wall intersecting their direct line of sight). These unobstructed connections are presented in Figure 4.23, where the connection distances are implied by their colors (green for short connections and red for further connections). Since longer connections in an indoor environment have an increased probability that some wall will obstruct them, it is reasonable that the identified unobstructed connections are mostly short ones.

The representative attenuation of each unobstructed Np-Sp pair was plotted against the physical distance between them (i.e. the connection distance), producing the graph of Figure 4.24. These values were then used to develop a Free Space Path Loss (FSPL) model based on which, attenuation estimations may be produced given some distance. As suggested by this model, the attenuation of a BLE signal would still reach -100dBm after 60 meters of unobstructed propagation within the selected indoor environment. The reason why the value of -100dBm has been considered here, is because signals that are weaker than that tend to be missed by receivers. Therefore, as also suggested in subsection 4.2.1, RSSIs lower than -100dBm shall be generally ignored.

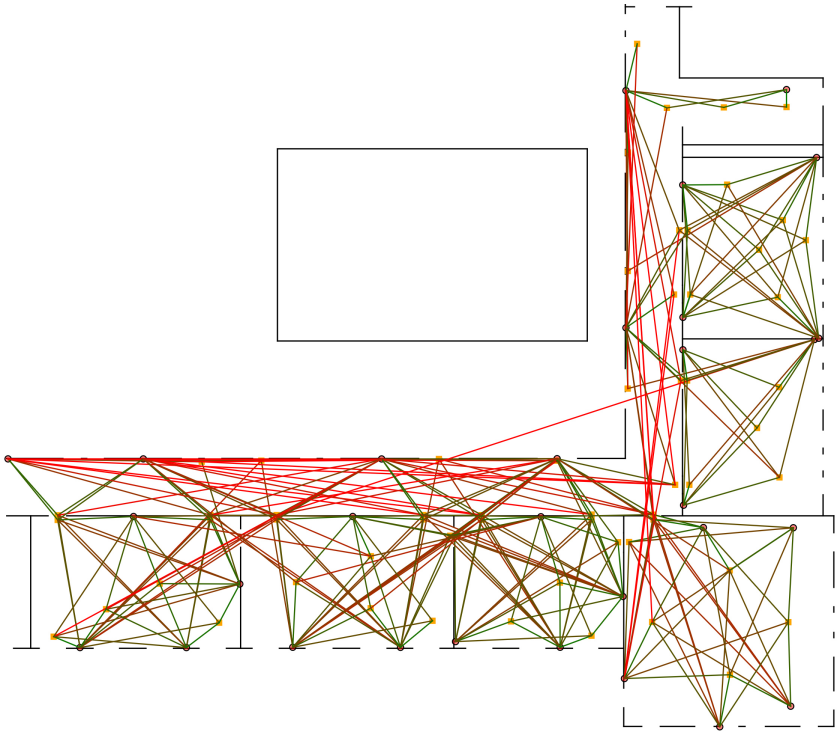


Figure 4.23: Unobstructed connections of Node/Sample Positions

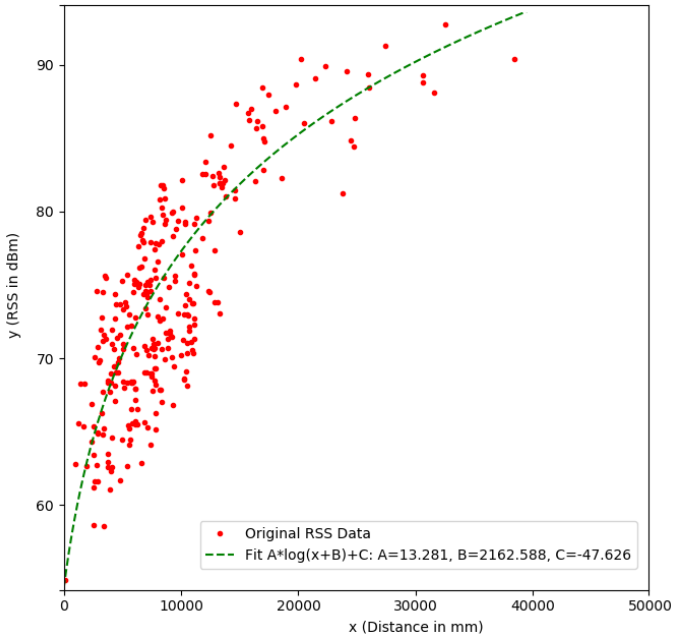


Figure 4.24: FSPL Distance to RSSI Observations

So far, we have a model which can be used to estimate at a position x that lies at some distance from the transmitting BLE node, the attenuation of an unobstructed signal (according to that propagation distance). However, signals may not always be able to reach that position x without being obstructed and/or reflected. For that reason, we need to develop also a model that shall handle these cases. As already specified, the attenuation of a signal at any position depends a) on the attenuation due to the travelled distance (described by the FSPL model we already developed) and b) the attenuation due to the number of reflections and refractions. Since in our indoor model, 3 different obstruction types have been used (Figure 4.18), the total amount of attenuation coefficients are 6 (Figure 4.25).

	Reflection Coefficient	***Refraction Coefficient***
Thick Concrete Wall	a	d
Thin Concrete Wall	b	e
Wooden Wall	c	f

Figure 4.25: Labels of the Attenuation Coefficients

To be able to compute the value of each coefficient and thus, enable the development of a propagation simulator that is able to handle also reflections and refractions, we utilized our indoor model to launch 1080 equally distributed rays from every node (Np). Each ray had a length of 60 meters and for that distance, we tracked the number of reflections and refractions for all possible trajectories.

To understand better this concept, we can refer to the Ray Launching Example of Figure 4.20. There, the Np is the Rays Launching Point and from the equally distributed (light blue) rays, it can be seen that only 2 manage in some way to reach their destination (purple Sp). In one case via a direct line of sight (originating from Ray1), and in another case via one reflection (originating from Ray2). The latter path is one of the three different trajectory scenarios that generally correspond to Ray2. The other scenarios would be a) 1 Refraction & 1 Reflection and b) 2 Refractions. From this description, it might not be immediately straightforward, yet for 30 nodes, the total amount of trajectories for which we had to ultimately track the corresponding amount of reflections and refractions, were billions. An example of trajectory scenarios corresponding to one node position can be seen in Figure 4.26.

To train our propagation model (i.e. to find the best coefficients), the 1800 representative attenuations were used as the ground truth through a continuous function of 4 iterated steps. The first one was to select some (logical) attenuation coefficients. The next one was to apply these coefficients to the billions of possible trajectory scenarios and to compute for each case, the final resulting attenuation.

Then, the next step was to identify for each of the 1800 Np-Sp connections, the trajectory offering the least attenuation (i.e. find the path from which the strongest signal came). The last step was to compute the total difference between the 1800 attenuations of the ground truth and the 1800 simulated attenuations. Finally, these 4 steps were repeated until the total difference was minimized and every time, at the first step, some new coefficients were evaluated. A process leading to the development of a radio propagation model that offered an average RSSI error of -3.39dBm (compared to the ground truth).

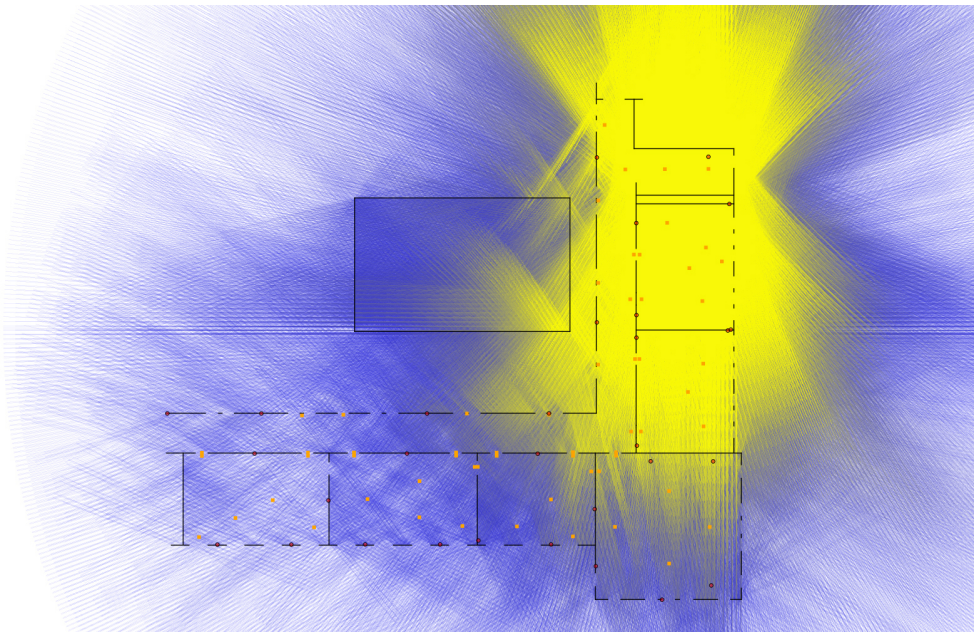


Figure 4.26: Training the Radio Propagation Engine

At this point, it should be stated that having not trained our propagation model and having used instead coefficients taken from literature research, the best achieved average RSSI error would be ≈ -11dBm. Moreover, our training methodology relies on the fact that the attenuation of a propagated signal does not depend on the energy of the incidental signal, as depicted in Figure 4.27.

The developed model is considered to be among the most crucial components for the whole optimization process. It enables us to quickly simulate different deployment scenarios via generated Radiomaps, like the one being presented in Figure 4.28. In this example, the attenuation of the signals being transmitted from the green node, is visualized, across the indoor environment.

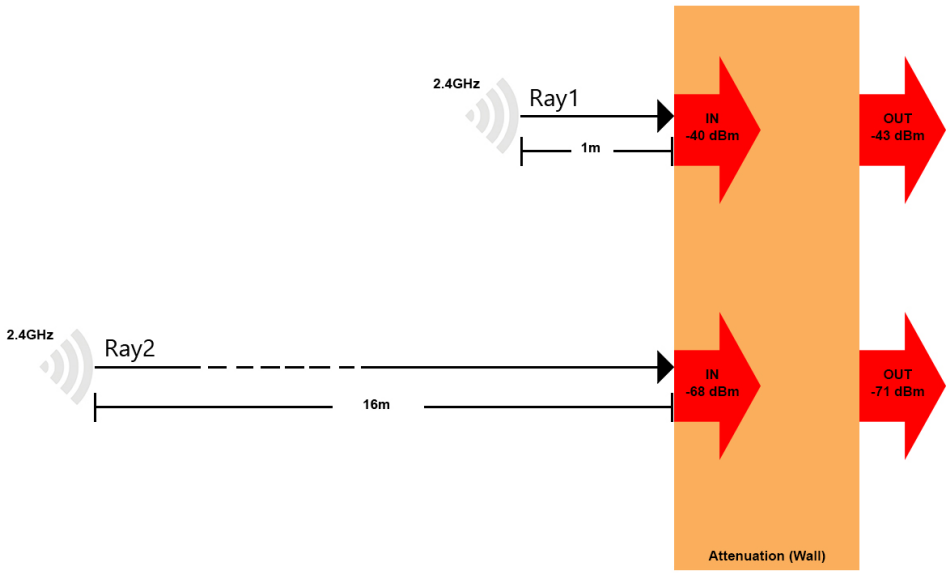


Figure 4.27: Radio Attenuation due to Refraction

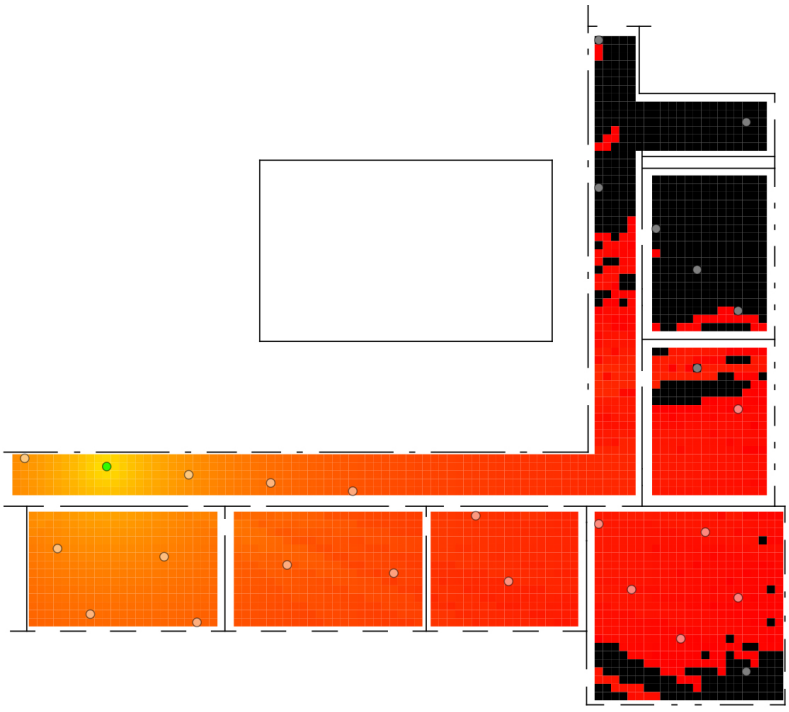


Figure 4.28: Radio Propagation Example from a BLE Node

4.5. Genetic Algorithm Integration

Having developed the simulation engine that will allow us to quickly assess different node-placement scenarios, the next step is to develop the algorithm that will handle the whole optimization process. This algorithm shall work as a black box where the input data will be 1) the 2D model of the indoor environment and 2) the number of nodes needed to be deployed within it in an optimal way, according to how this was defined in Section 4.2.5 and Section 4.2.6. Eventually, the output of this black box shall be the optimal positions of these nodes, along with the corresponding Radiomap which can be used to evaluate the solution.

During the optimization process it is not feasible to assess every possible deployment scenario. Nonetheless, there are alternatives which can be borrowed from the field of Artificial Intelligence, that can help us converge quickly to a solution being close to the optimal one. Such an approach is the use of a Genetic Algorithm, the implementation of which shall be discussed in this chapter.

4.5.1. Data Preparation

The model of the indoor environment that will be used during the optimization needs to respect the rules explained in Section 4.3 regarding the definition of the zones (i.e. the locations that we are interested in making more distinct within an Indoor Positioning System). Using that model and by following the methodology proposed in Section 4.2.4, the separation distances can be generated. For each cell that is part of this connectivity, it is required that we know for every possible node position, the corresponding received attenuation (RSSI). Creating such an RSSI mapping will purge the need of re-executing the expensive ray-launching process every time we do a check. Figure 4.29 illustrates an example of such an RSSI-mapping. Allowing nodes to be deployed only at the zone borders (i.e. assuming these represent walls), then in this figure, the green circle represents such a possible node position. Then, the attenuation that a separation cell would have from that node is presented using the red-yellow color-scheme.

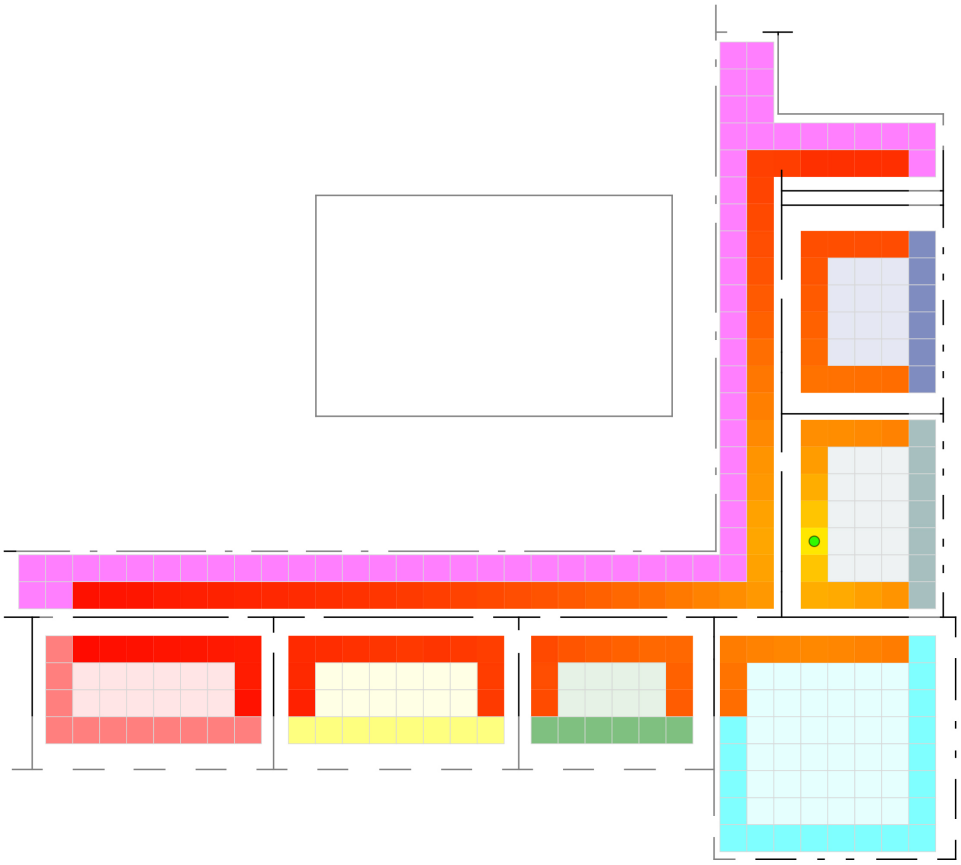


Figure 4.29: RSSIs Mapping for a Node Position

4.5.2. The Encoding

The utilization of a Genetic Algorithm for solving a problem requires that the problem is first encoded in accordance with the GA's life circle that was presented in Section 3.3.1. Since every implementation is unique, this section presents a new one that specifically applies to the purposes of our optimization's problem.

In a Genetic Algorithm, a possible solution (e.g. the deployment positions of the nodes) is represented by an Individual (Chromosome). This Individual is composed from a sequence of genes, each one of which has often a binary (0/1) identity. In our case, the size of this sequence of genes can reflect the possible positions where a node can be installed. Every different gene, or cell, may have a binary state depending on whether a node is installed there (1) or not (0). Therefore, arranging all cells in such a sequence, leads to the creation of an Individual. An example of such a Chromosome is presented in Figure 4.30 where 5 nodes have been placed in 5 specific positions. Since during the optimization, the setup size is constant, the amount of the allowed "non-zero" genes is also constant.

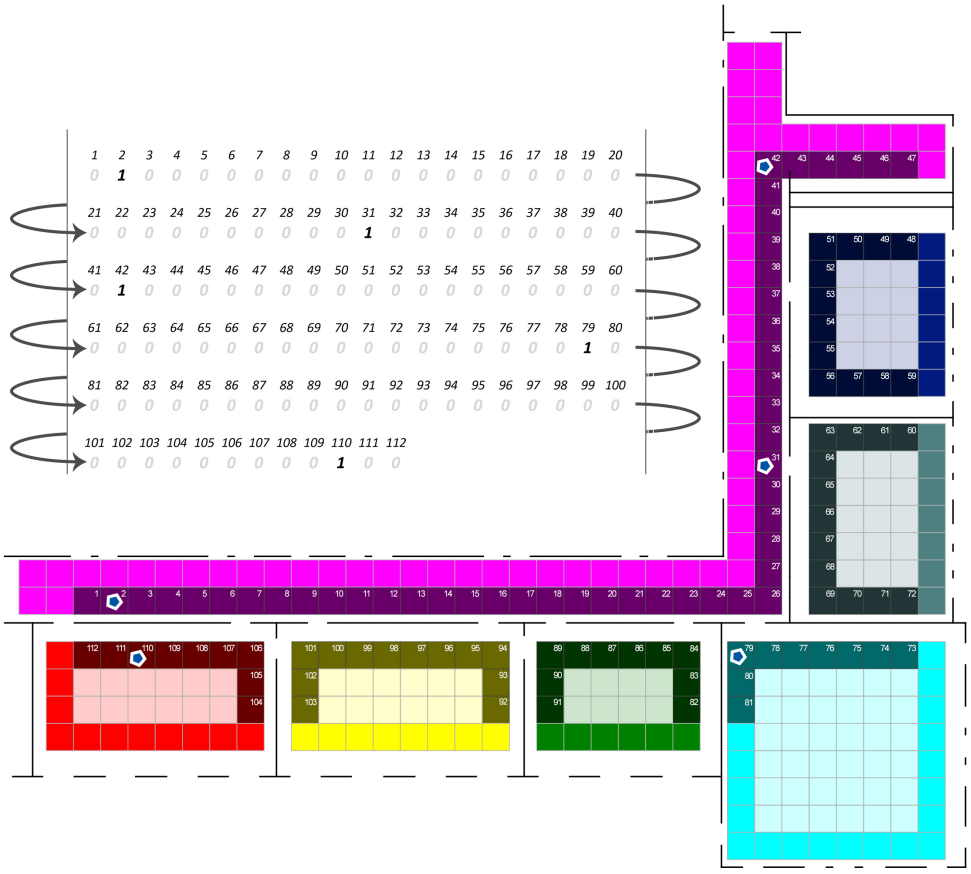


Figure 4.30: An Individual Chromosome (or Solution) for a 5-Node Setup

It is evident that the indoor model used in this chapter's examples has a cell size which is much bigger than the ones used in the optimization's test cases. This has led to the generation of a relatively small number of possible placement positions, when the same number in the first test case is 840 and thus, every Individual there is composed by a sequence of 840 genes.

The very first stage of a GA is the generation of the initial Population of random Individuals, the amount of which typically remains the same throughout the generations successions. After assessing different Population sizes in terms of their converging performance, the one that was used in the end was 105. Therefore, at the first stage, each one of these 105 Individuals had a random gene sequence, respecting always the required amount of the "non-zero" genes (according to the setup size). This Population shall, via the iterative process of natural selection, lead to better and better generations of solutions.

After the generation of the initial Population of Individuals (i.e. solutions), a repetitive circle begins involving 1) fitness checks, based on which, the fittest Individuals are selected and 2) crossovers of these Individuals leading to new and strongest generations of the Population. Although it is possible to introduce a convergence threshold which can be used to trigger the termination of this circle, our implementation avoids doing so because it is difficult to do quantitative predictions of the performance metrics that are used to measure the localization distinctiveness.

4.5.3. Fitness Check and Selection

In Section 4.2.5 and Section 4.2.6 2 different metrics were developed for measuring the localization performance of an Indoor Positioning System. For each one of these metrics, an assessment method has been developed, taking as input, the Individual solution and outputting the corresponding measured performance. It should be stated that during a single optimization process, only one of these methods is considered.

During the Selection phase, the employed assessment method is used to measure the performance of each one of the 105 Individuals within the Population. Then, the 50 fittest Individuals (i.e. deployment scenarios offering the best localization performance according to the used metric) are selected and combined in pairs. This results in 25 pairs of parents, where each pair shall breed 4 new Individuals.

4.5.4. Crossover and Mutation

Having identified the 25 strongest pairs of Chromosomes that will become parents, we proceed to the crossover & mutation phase, from which the new generation shall emerge, replacing the previous one. To begin with, the strongest Individual which is among the parents, is directly copied to the new generation to ensure that the fittest solution will not be lost. Along with it, 4 new random Individuals are additionally added to the next generation to ensure that the optimization algorithm spends also some resources for actively searching the vast space of solutions.

As mentioned, every pair from the 25 selected ones shall generate 4 new Chromosomes that each will inherit some characteristics from each parent. Hence, 100 new Chromosomes along with the 5 previously mentioned ones which directly passed to the next generation, will compose the new Population of 105 again Individuals. The exact process is depicted in Figure 4.31, where a simplified example of the genesis of a new Individual from 2 parents is showed.

In this example, each parent has 10 genes, 4 of which have a "non-zero" identity (i.e. a node is installed at that cell). The first step of the crossover phase is to identify which genes between the parents are not identical. In these cases, a random gene selection is made, leading eventually to a new sequence of genes with probably inherited characteristics from both parents. Then, a mutation is performed

during which, every gene has a small probability to switch its identity. It was observed that this probability should depend on the size of the sequence because too many mutations result in a very slow convergence. In the example, the purple genes are the ones that received a mutation, although this change corresponds to a probability of 2/10 which is an exaggeration. After the mutation, the new generated Chromosome will have probably an invalid amount of "non-zero" genes. Therefore, a validation check is required that will randomly correct the Individual and convert it into a valid one. After the reproduction of the new generation, one circle of natural selection has closed and the next begins. This involves a brand new Selection and Crossover process, the continuation of which is allowed until a convergence has been observed.

The total amount of cells along with the setup size is critical to the complexity of the algorithm since these, together, define the number of all possible combinations. These can be calculated using the expression:

$$\frac{n!}{p!(n-p)!}$$

where n the nodes in the setup and p the available positions. Hence, the amount of all possible combinations for the 30-nodes optimization on the first test case is $\approx$ 1.2e-55.

This 55-digit number makes it clear that the search space where the optimal solution lies, is vast. For that reason, the more computational power is spent for this search, the better and so, designing the Genetic Algorithm to be able to utilize more computational resources is highly beneficial. With that in mind, the exact implementation of our Genetic Algorithm supports parallel threads which can communicate with each other when a thread has identified a better global Individual. Moreover, each thread has a different mutation rate which allows for a wide search variety; from a highly random search, to fine-searching near local optima.

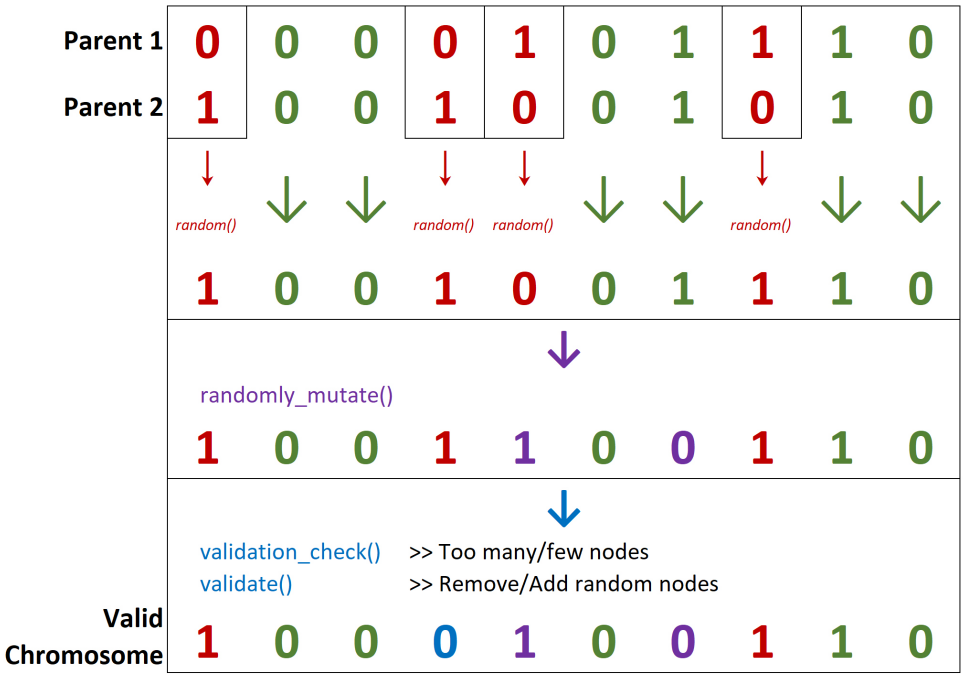


Figure 4.31: The Phases of Crossover and Mutation
(based on a 4-Node Setup)

4.5.5. Genetic Algorithm Challenges

Although the performance of our GA's implementation in terms of solution quality shall be discussed in detail in the next chapter (Chapter 6), some important considerations regarding its speed performance should be noted.

To begin with, after finding a new optimal solution, then, the more time is spend on trying to find an even better one, the higher the probability that the algorithm spends resources for assessing solutions that have already been reassessed. That, since the vast number of possible solutions do not allow for tracking which Chromosomes have already been generated (and thus, become possible to prevent their re-generation).

Furthermore, after having stuck into a local optimum and assuming that the true best solution has not already been found, the algorithm relies mostly on the mutation mechanism (and in our case, the random Individuals too) to be able to escape it. However, this escaping capability is inversely proportional to the capability of convergence towards a finest solution, making the selection of the ideal threshold a notable challenge. Experiments showed that within the first 15 minutes and for a 30-nodes setup, our algorithm has typically converged to an optimal solution which is very similar to the optimal solution of other optimization instances.

4.6. Improving Further the Searching Functionality

Decreasing the computational calculations is crucial for the optimization. For that, several practices can be followed which are discussed below.

In Section 4.2, it became evident that the continuous space (Figure 4.32) needs to be converted into a discrete one, for assigning RSSI values across the cells. This will generate a grid of cells and since its resolution (e.g. Figure 4.33 or Figure 4.34) has a direct impact on both the computational load and the accuracy, its cell-size shall be parameterizable to be able to explore different accuracy/complexity ratios.

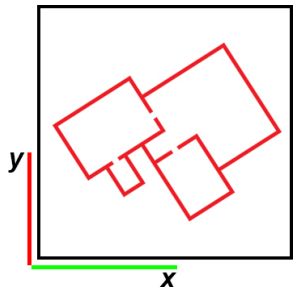


Figure 4.32: Continuous space

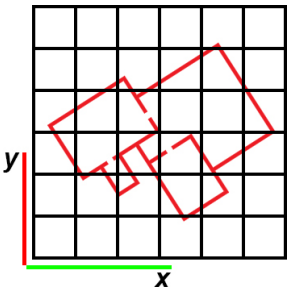


Figure 4.33: Large cell size

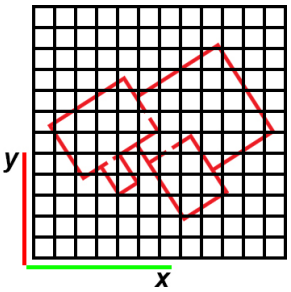


Figure 4.34: 2x smaller cell size

With that in mind, two different grid resolutions will be used for comparison purposes; one based on a cell-size of 45cm and another one based on a cell-size of 72cm. For the first case and in accordance with the methodology explained in Section 4.2.4 regarding the modeling of the separation distances, the corresponding connections are presented in Figure 4.35, whereas for the second case (i.e. 72cm cell-size), the corresponding connections are shown in Figure 4.36.

An additional way for improving the search functionality is to ensure perpendicularity and parallelism for the building model. As it has already been mentioned, an excessive number of signal rays is expected to be launched for intersection checks against various wall features. For that reason, and since most building models always maintain a 90° angle perpendicularity, we can significantly reduce the geometrical calculations by ensuring that each feature in the building model is perpendicular or parallel to the X, Y reference axes.

In the end, the computational complexity of the optimization is expected to be mostly affected by the amount of the nodes and the size of the area. Therefore, if the entire process could be divided, based on those factors, into many different parallel operations, it would enable us to utilize the power of parallel computing and improve further the search performance.

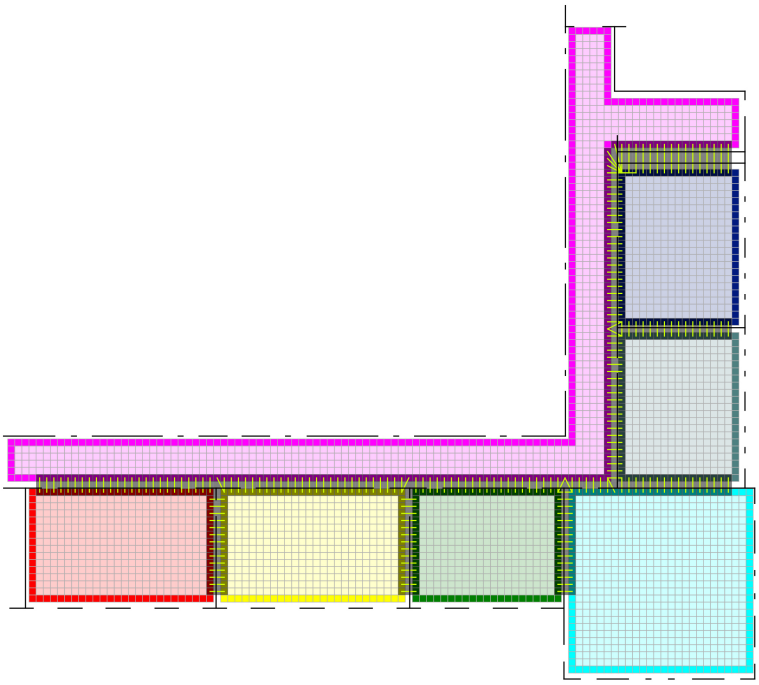


Figure 4.35: Separation Distances of the Zones at 45cm Cell-Size

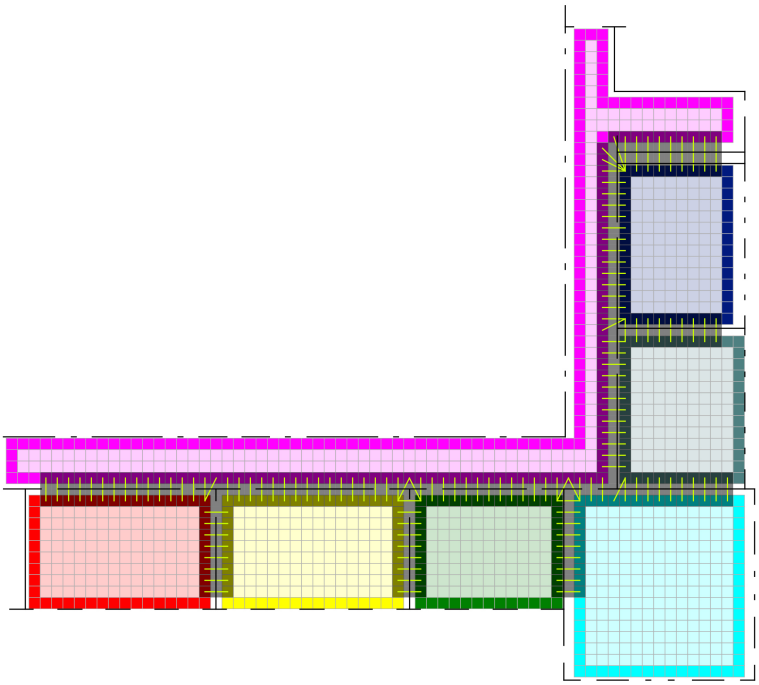


Figure 4.36: Separation Distances of the Zones at 72cm Cell-Size

5

Evaluation Process

In Chapter 4, a discussion was made regarding each component that participates in the whole optimization process. This process has been integrated in "Mesh Optimizer"; a software developed during this graduation project which can identify node deployments of any setup size that offer improved localization performance. However, being aware of the extent of this performance gain is quite important and so, the optimization algorithm needs to be also evaluated. Therefore, this chapter describes the entire evaluation's setup and process.

To begin with, the evaluation shall be performed considering the environment of our case study which is within the faculty of Architecture and the Built Environment at TU Delft (Section 1.3). For this evaluation, the localization performance of 3 different (in terms of setup size) non-optimized deployment scenarios will be compared with the localization performance of the corresponding optimized ones. The 3 non-optimized setups are shown in Figure 5.1 (30-nodes setup), Figure 5.2 (15-nodes setup) and Figure 5.3 (5-nodes setup). The node deployment in each non-optimized scenario respects common practices that administrators often follow during IPS installations (i.e. regular placement).

To compute the localization performance in each deployment scenario, the locations of 60 sample cells shall be classified and compared with the ground truth. These cells are presented in Figure 5.4 along with their IDs and true locations (the ground truth). Hence, this figure can be used as a reference for mapping between zones and cells. It should be also mentioned that throughout this evaluation, the solutions for the grid of the highest resolution (i.e. 45cm) shall be used.

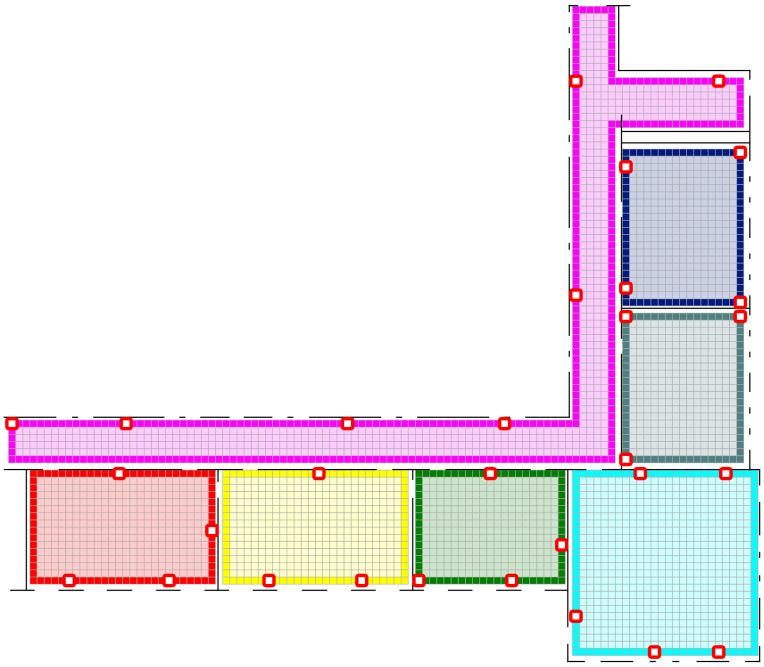


Figure 5.1: None-Optimized Regular (Regular) Deployment of 30 Nodes

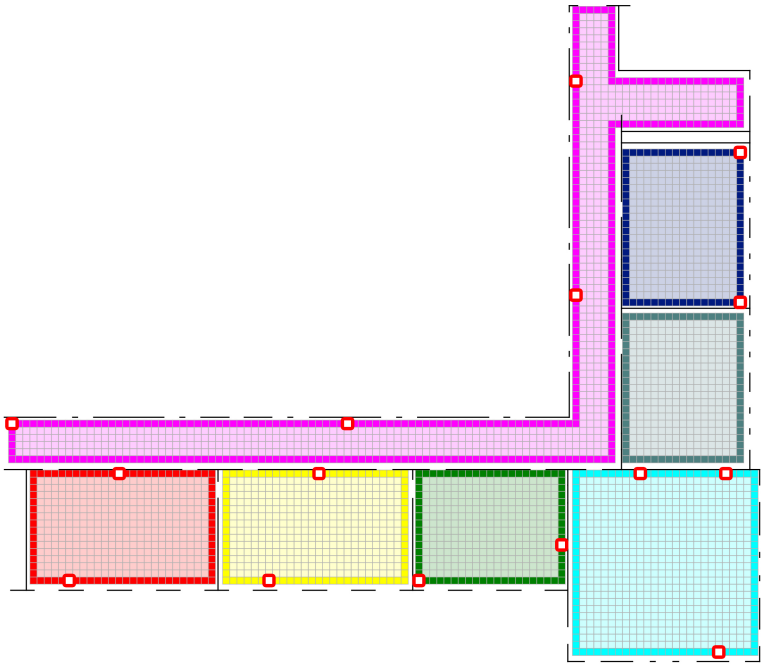


Figure 5.2: None-Optimized (Regular) Deployment of 15 Nodes

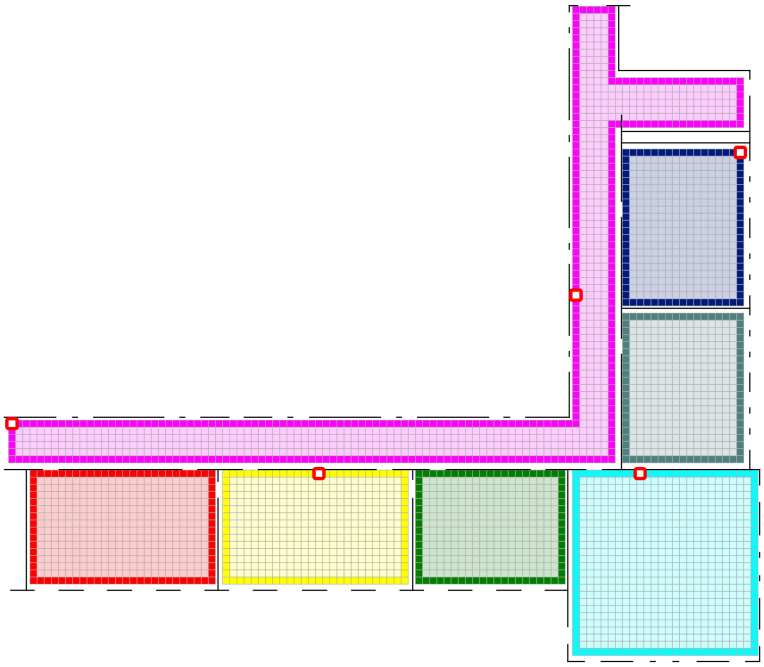


Figure 5.3: None-Optimized (Regular) Deployment of 5 Nodes

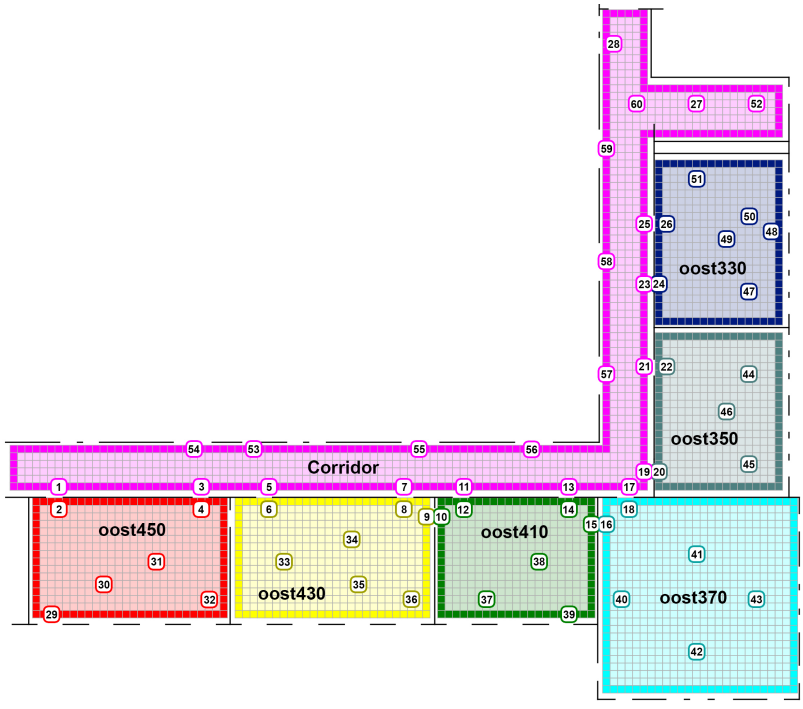


Figure 5.4: Sample Cells used for the Simulated Evaluation

One could notice that the 60 sample cells which have been selected for the evaluation, match the sample positions that were utilized during the RSSI gathering that was used to train our radio simulator. Moreover, one can also notice that the non-optimized deployments match the node deployment (Figure 5.5) that was apparent during that RSSI gathering. This means that in case we wished to evaluate in real life our 3 optimized solutions, we would not have to sample again for the non-optimized scenarios (which, alone, is 3 days of work-load). Lastly, it should be marked that although these sample positions are well spread across the area (considering also all door openings), the simulated evaluation might ideally consider not only 60 cells, but all of them.

The localization algorithm that will be used for the classification of (the location of) each sample point, is the kNN algorithm. According to that, the k nearest (in Radio Signature units) neighbor-cells are selected and their locations are used to identify which location has been detected most of the times. Although the most common kappa coefficient for positioning purposes is 1, two others (3 and 5) shall be additionally used to enable comparisons. Using 1 as kappa would always return a correct localization because a) the set of the evaluated cells is the same as the set of the neighbor-cells (being considered for the localization) and b) there is no RS variation in the system. Therefore, the closest neighbor of a cell being evaluated, is always the cell itself at zero distance.

Since the most often kappa coefficient is 1, another approach shall be used to assess its expected performance. Namely, the distance to the first wrong neighbor-cell. The term wrong is used here to denote that the neighbor-cell belongs to a zone that is not the same as the zone where the evaluated cell belongs. This will suggest how close (in RS units) we are at that position, to a wrong estimation.

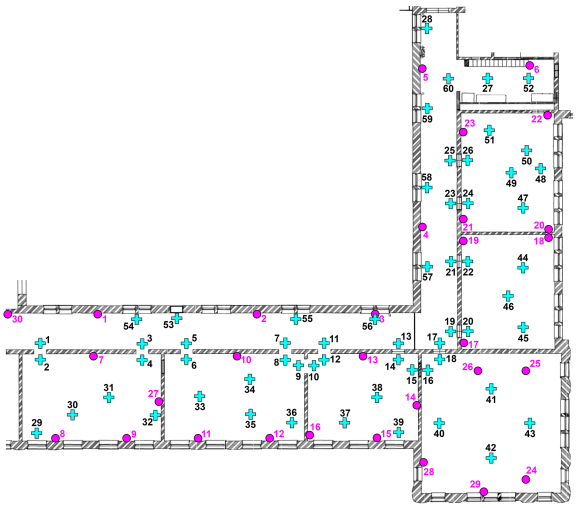


Figure 5.5: Sample & Node Positions for the RSSI gathering

6

Results & Interpretation

6.1. Deployment Solutions Found

6.1.1. Setups for an Optimal Minimum Separation Distance

This section presents the optimal node setups that the proposed optimization discovered while considering the first performance metric (i.e. minimum separation distance). In each case and as mentioned in Section 4.6, two different grid resolutions are used for comparison purposes; one based on a cell-size of 45cm and another one based on a cell-size of 72cm.

For a 30-nodes setup, the optimal node placements are presented in Figure 6.1 and Figure 6.2 accordingly. From the perspective of human's intelligence, it can be noticed that positions where the signals can pass from a zone to the other while undergoing low attenuation, are highly favoured. Such places are the ones close to door openings (Figure 6.3) or windows. Moreover, most of the placement positions lie within the separation area, whereas some few can also be found at the far ends of the zones as presented in Figure 6.4. A comparison between the two different cell-sizes suggests that the setups do not match exactly. This may be related to the fact that the "grid-ification" process of the zones produced a) zone borders that are not identical and b) cells of different geometries and thus, different interaction characteristics with the propagated rays of the radio simulator. Yet, in most of the regions, one can identify common deployment patterns.

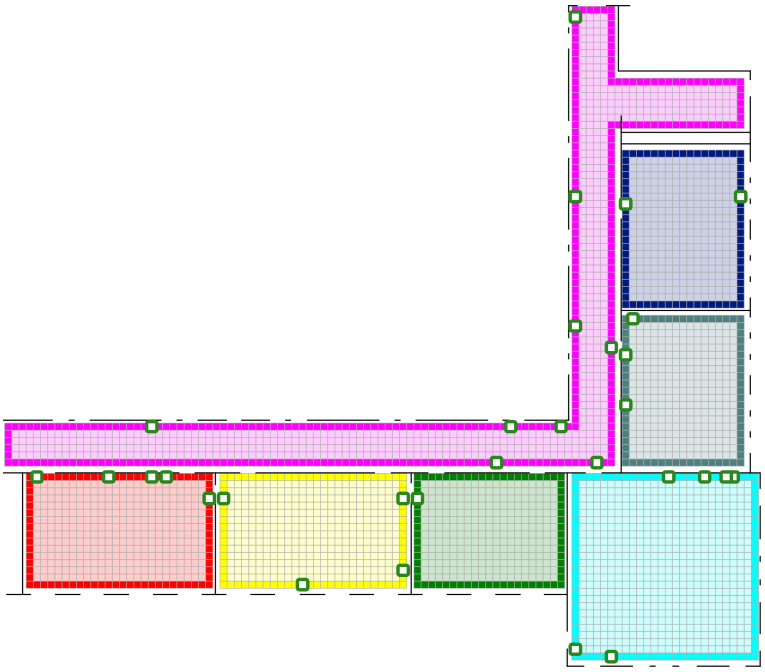


Figure 6.1: Optimal Deployment for 30 Nodes (Opt1 - Cell Size 45cm)

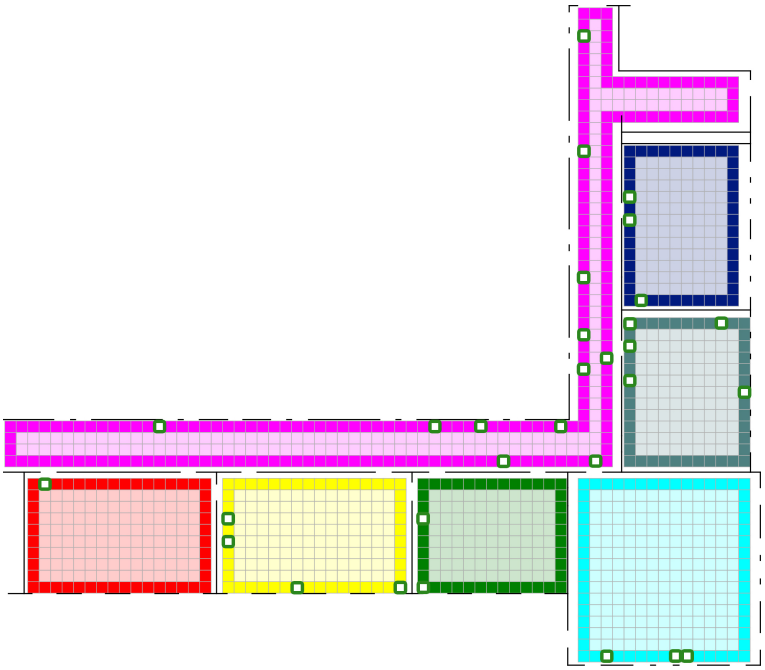


Figure 6.2: Optimal Deployment for 30 Nodes (Opt1 - Cell Size 72cm)

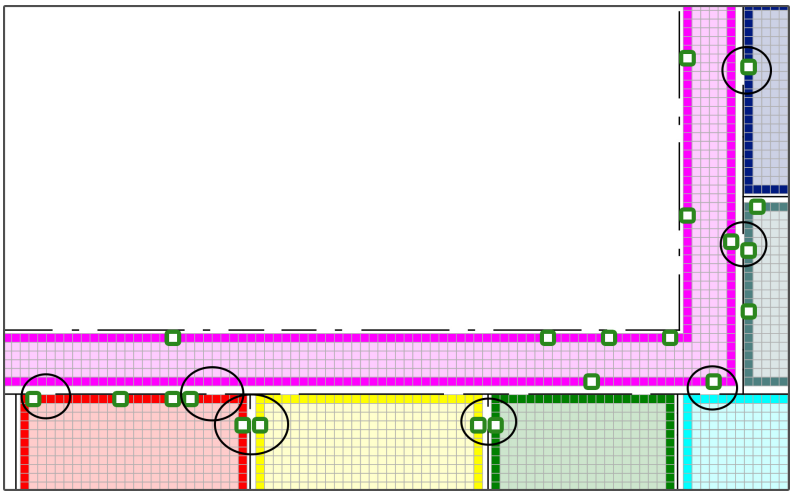


Figure 6.3: Node Placement next to Door Openings

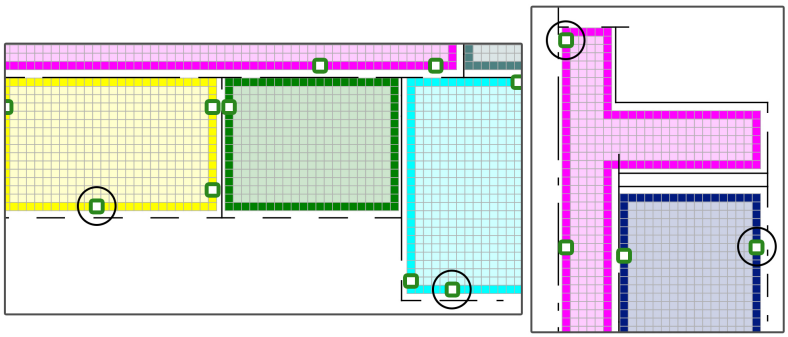


Figure 6.4: Node Placement at the far Side of Zones

For a 15-nodes setup, the optimal deployments are presented in Figure 6.5 and Figure 6.6 accordingly. As in the previous case, although the placement between the different cell-sizes does respect common patterns, it is not identical. Contrariwise, when comparing each deployment solution with the corresponding bigger setup above (i.e. 30 nodes), several cell positions can be even found to be exactly the same. Also, once again, positions next to obstruction-less zone borders seem to have been favoured at some level. Lastly, the best solutions that the optimization discovered for a 5-nodes setup are presented in Figure 6.7 and Figure 6.8 accordingly. Although the aforementioned observations are applicable in these cases too, it should further be noticed that the deployment of the grid with cell-size 45cm (Figure 6.7) is almost a subset of the placement presented in Figure 6.6 (15 Nodes, Cell-Size 72cm optimization).

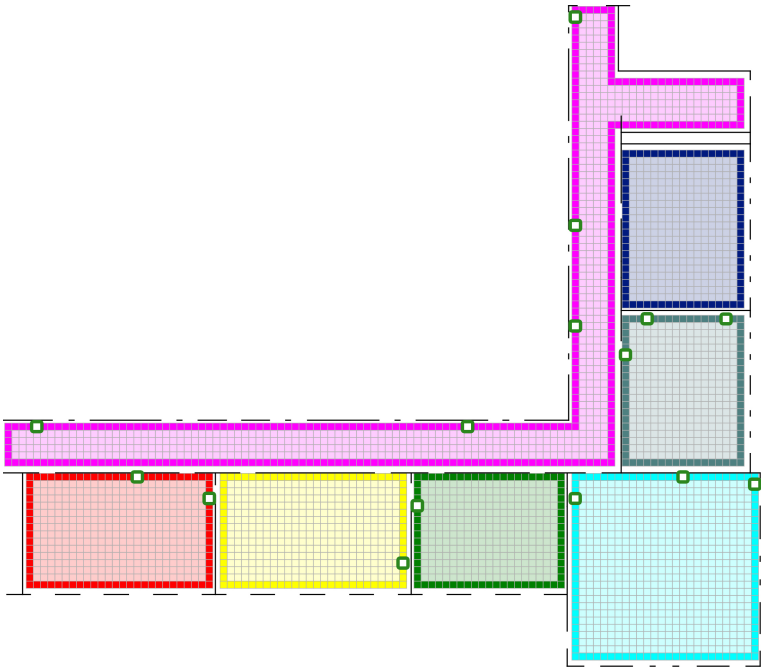


Figure 6.5: Optimal Deployment for 15 Nodes (Opt1 - Cell Size 45cm)

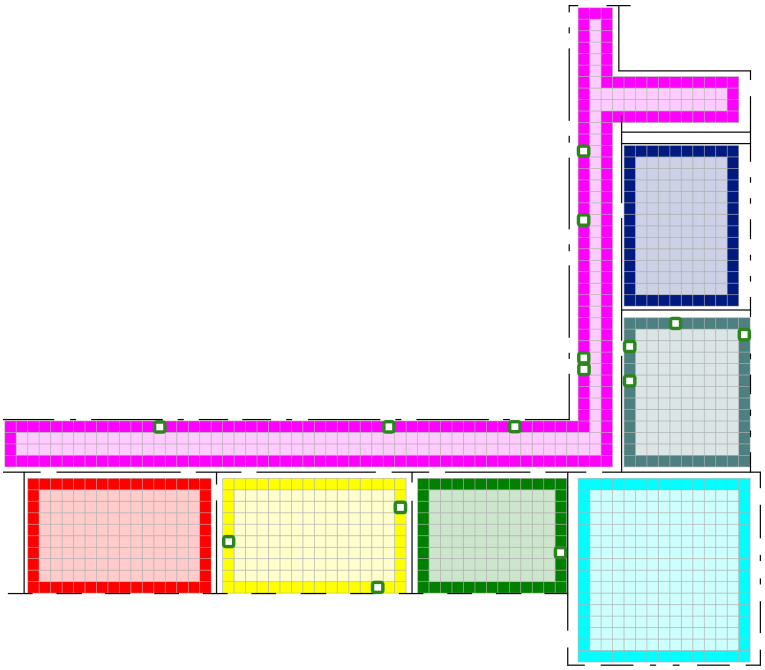


Figure 6.6: Optimal Deployment for 15 Nodes (Opt1 - Cell Size 72cm)

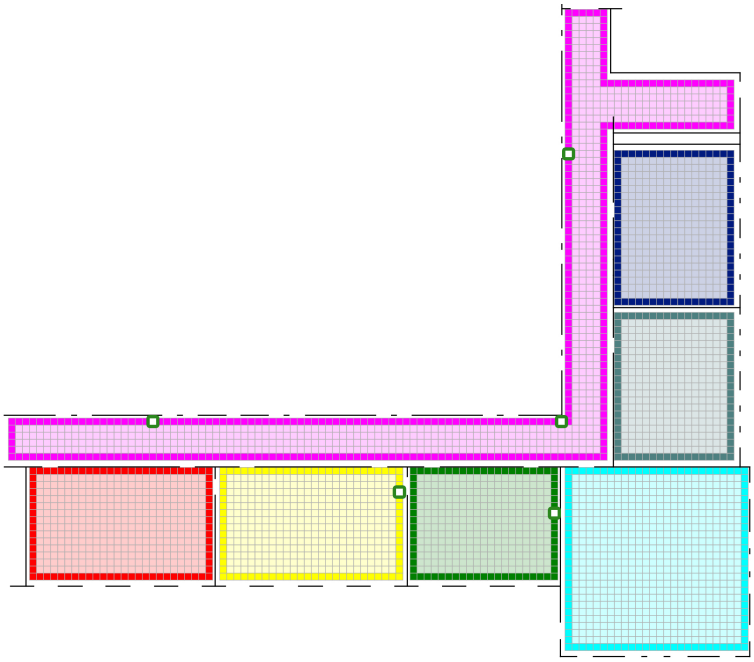


Figure 6.7: Optimal Deployment for 5 Nodes (Opt1 - Cell Size 45cm)

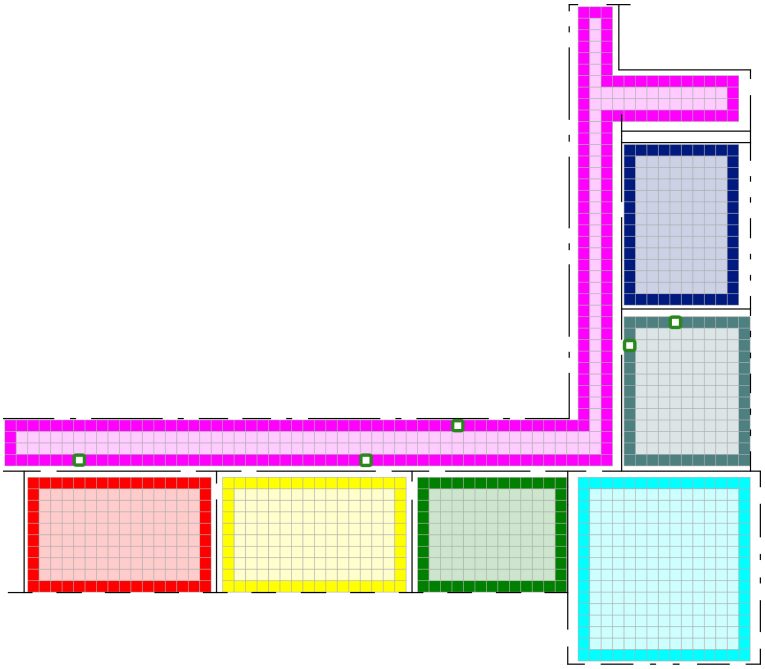


Figure 6.8: Optimal Deployment for 5 Nodes (Opt1 - Cell Size 72cm)

6.1.2. Setups for an Optimal Product of n Shortest Separation Distances

This section presents the optimal solutions (Figure 6.9 - Figure 6.11) that our optimization discovered when considering the second proposed performance metric; namely, the product of the n shortest separation distances. The maximization of this metric was driven again by the Genetic Algorithm, but in this case, only the most detailed grid resolution has been assessed (i.e. Cell Size of 45cm), whereas for the coefficient n, the percentage of 5% has been selected. These shall also be the deployments that will be evaluated in the following sections.

As explained in subsection 4.2.6, the objective of this 2nd optimization is to attempt to increase further the overall separation distances and thus, lead hopefully to better localization performance. However, in the following figures, the impact of this new objective cannot easily be noticed. In general, the same patterns as in the previous optimizations can be found, leading to no clear, intuitively, indication of what the differences between the deployment scenarios are.

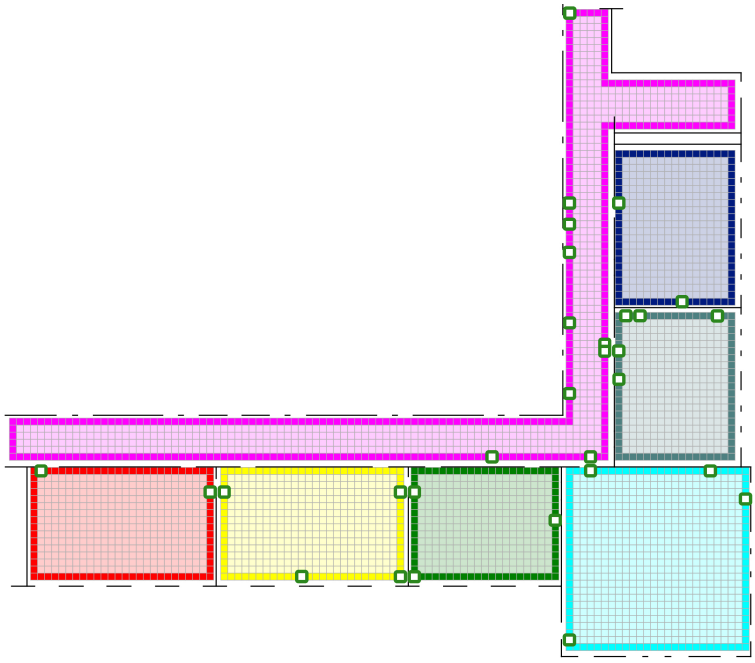


Figure 6.9: Optimal Deployment for 30 Nodes (Opt2 - Cell Size 45cm)

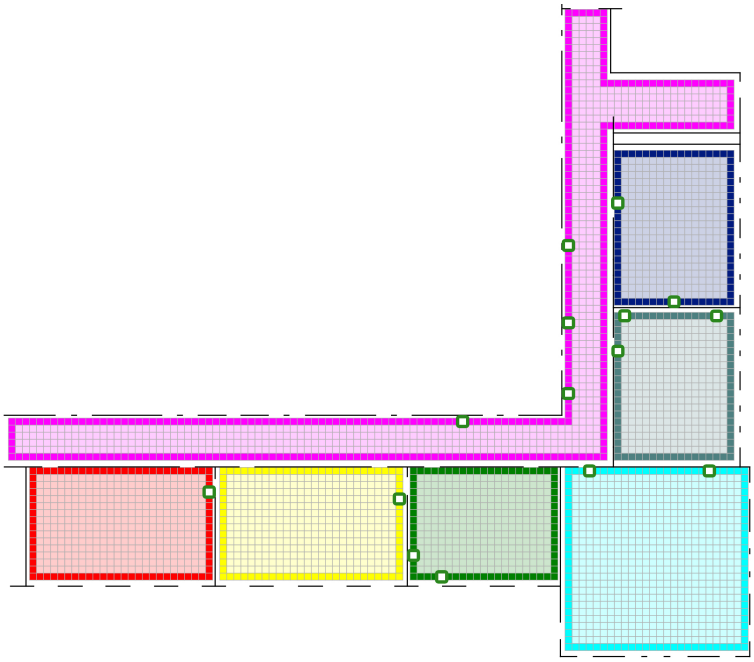


Figure 6.10: Optimal Deployment for 15 Nodes (Opt2 - Cell Size 72cm)

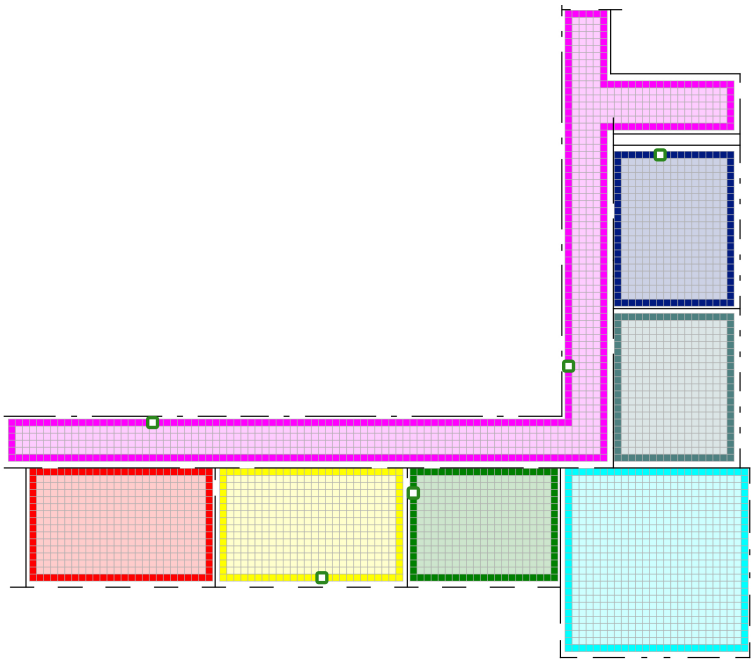


Figure 6.11: Optimal Deployment for 5 Nodes (Opt2 - Cell Size 45cm)

6.2. Evaluation of the Localization Improvement

Respecting the evaluation's process presented in Chapter 5, the next step is to assess in terms of localization performance, our optimal deployment solutions against the corresponding non-optimal scenarios.

6.2.1. Assessing the kNN performance for k=1

Comparing the localization performance between the non-optimized and the optimized scenarios results in the data of Figure 6.12 and Figure 6.13. The first figure corresponds to the deployment solutions that were produced via the optimization of the minimum separation distance, whereas the second figure corresponds to the solutions that were produced via the optimization of the product of the 5% shortest separation distances. The second figure includes also the differences between these two metrics.

Starting with the first case (minimum separation distance in Figure 6.12), one can notice 3 different tables. Each table represents a different node setup size, starting at the left with the biggest one (30 nodes) and ending at the right with the smallest one (5 nodes). Every table has 5 columns. The first column contains the position IDs of the sample cells that are being evaluated. Then, each one of these cells belongs (in reality) to the zone that is shown within the second column. The third column corresponds to the non-optimized deployment and presents the distance in Radio Signature units to the closest wrong (i.e. belonging to another zone) neighbor-cell. Contrariwise, the fourth column presents the same feature, but within the optimized deployment. Lastly, the fifth column presents the localization improvement (in percentage) that the optimized deployment introduced. The higher this value is, the less probable for the localization to be incorrect.

Each table has been sorted based on its third column (i.e. the separation distances in the non-optimized deployments), resulting in a linear distance increase. A color gradient has also been used for the distances to make it easier to interpret the results. Moreover, the lowest and highest values within the 3rd and 4th columns are highlighted using colored bounding boxes.

The data of Figure 6.12 suggest an overall localization improvement which becomes more apparent as the node-setup decreases. Due to the aforementioned sorting, the cells where the localization is more prone to errors are found at the upper part of the table. There, the lower distances are represented with stronger shades of redness. Yet, a comparison between the upper part of the non-optimized and optimized columns suggests that in the optimized cases, a performance increase has been achieved since there, this shade becomes less intense. This improvement is notably apparent in the upper part of the 5-node setup optimization, where the separation improvements are reaching up to +237% (at sample position 9 which is the most problematic region in the non-optimized scenario).

30 Nodes				
Sample Pos	Zone	Non Opt	Opt	% Gain
25	*corridor*	6,8	12,0	75
26	*oost330*	6,8	12,0	75
13	*corridor*	7,4	8,2	11
14	*oost410*	7,4	8,2	11
5	*corridor*	8,5	9,8	16
6	*oost430*	8,5	9,8	16
9	*oost430*	8,5	11,0	29
10	*oost410*	8,5	11,0	29
1	*corridor*	8,6	9,8	14
2	*oost450*	8,6	9,8	14
11	*corridor*	9,2	11,1	21
12	*oost410*	9,2	11,1	21
3	*corridor*	9,4	10,9	16
4	*oost450*	9,4	10,9	16
17	*corridor*	9,4	11,7	24
18	*oost370*	9,4	11,7	24
7	*corridor*	9,8	7,8	-21
8	*oost430*	9,8	7,8	-21
19	*corridor*	10,6	11,5	8
20	*oost350*	10,6	11,5	8
21	*corridor*	10,8	12,2	12
22	*oost350*	10,8	12,2	12
23	*corridor*	11,8	10,6	-10
24	*oost330*	11,8	10,6	-10
15	*oost410*	11,9	12,0	1
16	*oost370*	11,9	12,0	1
53	*corridor*	13,8	17,1	24
51	*oost330*	16,3	20,1	23
54	*corridor*	16,4	19,5	19
49	*oost330*	17,8	20,3	14
33	*oost430*	18,2	18,5	1
55	*corridor*	18,7	20,6	10
58	*corridor*	18,8	15,5	-18
31	*oost450*	19,0	20,7	9
50	*oost330*	19,8	26,2	32
40	*oost370*	20,2	18,3	-9
57	*corridor*	20,4	26,8	31
30	*oost450*	21,0	17,9	-15
34	*oost430*	21,2	20,3	-4
59	*corridor*	21,4	21,1	-2
46	*oost350*	21,6	29,6	37
48	*oost330*	21,9	29,3	34
44	*oost350*	22,4	26,1	16
47	*oost330*	22,4	26,1	16
37	*oost410*	22,6	20,8	-8
56	*corridor*	22,7	18,8	-17
38	*oost410*	22,7	19,9	-12
35	*oost430*	23,6	25,3	7
41	*oost370*	23,8	20,5	-14
39	*oost410*	24,2	18,3	-24
32	*oost450*	24,4	23,6	-3
29	*oost450*	25,0	20,9	-17
45	*oost350*	25,5	20,5	-20
36	*oost430*	25,8	23,9	-7
60	*corridor*	26,4	22,1	-16
27	*corridor*	26,7	25,7	-4
43	*oost370*	29,5	27,4	-7
52	*corridor*	32,2	29,8	-7
28	*corridor*	32,9	34,1	4
42	*oost370*	40,1	29,0	-28
				7,28
				avg

15 Nodes				
Sample Pos	Zone	Non Opt	Opt	% Gain
9	*oost430*	4,0	6,7	70
10	*oost410*	4,0	6,7	70
25	*corridor*	4,7	6,2	32
26	*oost330*	4,7	6,2	32
13	*corridor*	5,1	5,0	-1
14	*oost410*	5,1	5,0	-1
1	*corridor*	5,3	7,3	37
2	*oost450*	5,3	7,3	37
7	*corridor*	5,4	6,7	23
8	*oost430*	5,4	6,7	23
11	*corridor*	6,5	7,7	19
12	*oost410*	6,5	7,7	19
5	*corridor*	6,7	8,6	29
6	*oost430*	6,7	8,6	29
3	*corridor*	6,9	7,1	3
4	*oost450*	6,9	7,1	3
17	*corridor*	7,4	7,2	-2
18	*oost370*	7,4	7,2	-2
23	*corridor*	7,9	8,3	5
24	*oost330*	7,9	8,3	5
19	*corridor*	8,1	8,0	-1
20	*oost350*	8,1	8,0	-1
16	*oost370*	8,4	10,2	22
15	*oost410*	8,6	10,2	18
53	*corridor*	9,1	10,9	21
21	*corridor*	9,4	9,0	-4
22	*oost350*	9,4	9,0	-4
40	*oost370*	10,0	13,4	35
54	*corridor*	10,3	13,9	35
58	*corridor*	11,0	14,8	34
51	*oost330*	11,1	11,8	6
56	*corridor*	12,0	15,7	30
55	*corridor*	12,2	16,2	32
31	*oost450*	12,3	13,9	13
39	*oost410*	12,5	13,4	7
38	*oost410*	12,6	14,3	13
34	*oost430*	12,7	14,2	11
49	*oost330*	12,8	15,1	18
57	*corridor*	13,6	18,8	39
33	*oost430*	13,7	14,5	5
41	*oost370*	13,8	15,6	13
32	*oost450*	14,3	14,4	1
36	*oost430*	14,5	18,3	27
37	*oost410*	15,1	14,7	-3
35	*oost430*	15,5	16,9	9
44	*oost350*	15,7	14,5	-8
47	*oost330*	15,7	14,5	-8
52	*corridor*	15,8	21,8	38
50	*oost330*	16,2	15,8	-3
30	*oost450*	16,2	16,7	3
46	*oost350*	16,4	21,9	34
59	*corridor*	16,5	11,8	-28
27	*corridor*	17,0	18,2	7
45	*oost350*	17,3	15,6	-10
48	*oost330*	17,8	19,1	7
43	*oost370*	20,3	19,5	-4
29	*oost450*	21,2	17,4	-18
60	*corridor*	21,5	16,2	-25
28	*corridor*	24,5	26,9	10
42	*oost370*	24,8	20,3	-18
				13,01
				avg

5 Nodes				
Sample Pos	Zone	Non Opt	Opt	% Gain
9	*oost430*	1,8	5,9	237
10	*oost410*	1,8	5,5	214
13	*corridor*	2,2	4,5	102
14	*oost410*	2,2	4,5	102
25	*corridor*	2,9	3,5	21
26	*oost330*	2,9	3,5	21
5	*corridor*	3,0	3,2	5
6	*oost430*	3,0	3,2	5
1	*corridor*	3,2	4,1	28
2	*oost450*	3,2	4,1	28
55	*corridor*	3,3	7,3	122
7	*corridor*	3,4	5,1	53
8	*oost430*	3,4	5,1	53
18	*oost370*	3,5	4,7	36
20	*oost350*	3,5	5,2	51
21	*corridor*	3,7	3,0	-21
22	*oost350*	3,7	3,0	-21
19	*corridor*	4,4	5,2	19
16	*oost370*	4,6	3,9	-15
17	*corridor*	4,6	4,7	2
23	*corridor*	4,6	3,5	-24
24	*oost330*	4,6	3,5	-24
11	*corridor*	4,7	4,3	-9
12	*oost410*	4,7	4,3	-9
3	*corridor*	4,9	3,3	-32
4	*oost450*	4,9	3,3	-32
31	*oost450*	4,9	5,7	17
35	*oost430*	5,0	8,8	76
56	*corridor*	5,7	8,6	52
53	*corridor*	5,9	5,3	-10
36	*oost430*	6,0	10,6	77
15	*oost410*	6,0	3,9	-36
40	*oost370*	6,0	6,0	0
32	*oost450*	6,3	9,3	47
38	*oost410*	6,4	9,4	47
57	*corridor*	6,4	8,4	31
33	*oost430*	6,4	6,3	-2
58	*corridor*	6,4	4,9	-24
54	*corridor*	6,6	8,2	25
37	*oost410*	6,9	9,1	31
34	*oost430*	7,1	7,1	0
47	*oost330*	7,3	8,2	13
59	*corridor*	7,3	12,3	68
39	*oost410*	7,4	6,0	-19
42	*oost370*	7,8	6,4	-18
41	*oost370*	8,1	7,4	-9
45	*oost350*	8,3	6,4	-23
43	*oost370*	8,4	6,7	-21
30	*oost450*	8,6	5,4	-37
49	*oost330*	8,9	8,8	0
51	*oost330*	8,9	5,0	-44
60	*corridor*	9,5	5,0	-48
48	*oost330*	9,7	10,2	5
52	*corridor*	9,7	9,2	-5
50	*oost330*	10,3	7,8	-24
46	*oost350*	10,4	9,0	-13
44	*oost350*	10,5	8,1	-23
27	*corridor*	10,9	8,2	-25
29	*oost450*	11,8	8,2	-30
28	*corridor*	14,0	7,8	-44
				15,77
				avg

Figure 6.12: Distances to Nearest Localization Error

This improvement, however, did not happen without a cost. It is evident that the performance at positions where the localization was exceptional (e.g. cell 42 which was 40,1 RS units away from the closest wrong neighbor) had to be reduced. This sounds intuitively correct since to gain performance somewhere, you may have to lose it from somewhere else. Nevertheless, in each case, the optimization led to a minimum separation distance that is higher compared to the one within non-optimized setups.

It is important to mention that although the optimization resulted (at some level) in a distance balancing between the upper and lower parts, in practice, this may not always mean more accurate localization. That, because, in reality, the accuracy of the localization depends (aside from the aforementioned distances) also from the noise of the Radio Signatures; a dependency which is unknown to us. To understand this notion better, we could assume that, due to noise, a correct estimation would demand a minimum distance of 40 units; a distance that was only achieved at cell 42 in the 30-nodes setup. However, by balancing the Radio Signatures and lowering this distance by -28%, would result in losing the only valuable position we already had.

The results of the optimization's assessment based on the 2nd performance metric (Product of n Shortest Distances), are shown in Figure 6.13. The structure of the presented tables is the same as in Figure 6.12, however, a new column has been added here showing the differences in performance gain between the two metrics. As in the previous case, one can notice again a general performance gain in the upper distances but also a loss in the lower ones.

On one hand, this metric showed that its purpose was fulfilled since, according to the last column, the optimization led to a bigger overall gain (i.e. in average: +2.2%, +4.6% and +6.9%) when compared to the performance gain of the 1st optimization. It should be noted that this increase was becoming bigger as the node-setup was decreasing. However, at the same time, the variance of these distances was also increasing which suggests a more intense change of the entire Radiomap. Lastly, it is important to emphasize that although the overall performance has been increased, to achieve that, the minimum distances had to become lower compared to the previous metric. Yet, these distances still remained bigger compared to the non-optimized scenario.

30 Nodes

Sample Pos	Zone	Non Opt	Opt	% Gain	*Metrics Gain Difference*
25	*corridor*	6,8	11,8	72	*-2,7*
26	*oost330*	6,8	11,8	72	*-2,7*
13	*corridor*	7,4	7,8	5	*-5,6*
14	*oost410*	7,4	7,8	5	*-5,6*
5	*corridor*	8,5	10,9	28	*12,6*
6	*oost430*	8,5	10,9	28	*12,6*
9	*oost430*	8,5	11,7	37	*7,9*
10	*oost410*	8,5	11,7	37	*7,9*
1	*corridor*	8,6	10,0	16	*2,4*
2	*oost450*	8,6	10,0	16	*2,4*
11	*corridor*	9,2	10,5	14	*-6,3*
12	*oost410*	9,2	10,5	14	*-6,3*
3	*corridor*	9,4	10,5	11	*-4,6*
4	*oost450*	9,4	10,5	11	*-4,6*
17	*corridor*	9,4	10,8	14	*-10,1*
18	*oost370*	9,4	10,8	14	*-10,1*
7	*corridor*	9,8	9,0	-9	*11,9*
8	*oost430*	9,8	9,0	-9	*11,9*
19	*corridor*	10,6	10,8	2	*-6,2*
20	*oost350*	10,6	10,8	2	*-6,2*
21	*corridor*	10,8	13,7	27	*14,4*
22	*oost350*	10,8	13,7	27	*14,4*
23	*corridor*	11,8	12,1	2	*12,4*
24	*oost330*	11,8	12,1	2	*12,4*
15	*oost410*	11,9	12,4	5	*3,8*
16	*oost370*	11,9	12,4	5	*3,8*
53	*corridor*	13,8	17,5	27	*2,8*
51	*oost330*	16,3	17,9	10	*-13,1*
54	*corridor*	16,4	18,4	12	*-6,6*
49	*oost330*	17,8	24,1	35	*21,3*
33	*oost430*	18,2	19,2	5	*4,0*
55	*corridor*	18,7	22,3	19	*9,2*
58	*corridor*	18,8	22,8	21	*39,1*
31	*oost450*	19,0	19,6	3	*-6,2*
50	*oost330*	19,8	22,3	13	*-19,8*
40	*oost370*	20,2	18,1	-10	*-1,1*
57	*corridor*	20,4	26,4	29	*-1,8*
30	*oost450*	21,0	19,3	-8	*6,5*
34	*oost430*	21,2	18,8	-11	*-7,2*
59	*corridor*	21,4	17,9	-16	*-14,7*
46	*oost350*	21,6	31,6	47	*9,4*
48	*oost330*	21,9	25,8	18	*-15,7*
44	*oost350*	22,4	23,5	5	*-11,4*
47	*oost330*	22,4	23,5	5	*-11,4*
37	*oost410*	22,6	21,8	-3	*4,6*
56	*corridor*	22,7	17,6	-22	*-5,2*
38	*oost410*	22,7	21,7	-5	*7,9*
35	*oost430*	23,6	26,8	13	*6,6*
41	*oost370*	23,8	24,7	4	*17,7*
39	*oost410*	24,2	18,1	-25	*-0,9*
32	*oost450*	24,4	20,6	-15	*-12,2*
29	*oost450*	25,0	20,3	-19	*-2,1*
45	*oost350*	25,5	24,7	-3	*16,6*
36	*oost430*	25,8	26,2	2	*8,9*
60	*corridor*	26,4	21,9	-17	*-0,4*
27	*corridor*	26,7	29,7	11	*14,8*
43	*oost370*	29,5	30,7	4	*11,2*
52	*corridor*	32,2	35,9	12	*18,9*
28	*corridor*	32,9	34,9	6	*2,2*
42	*oost370*	40,1	29,3	-27	*0,8*
				9,49 avg	***2,2*** avg

15 Nodes

Sample Pos	Zone	Non Opt	Opt	% Gain	*Metrics Gain Difference*
9	*oost430*	4,0	9,6	142	*71,4*
10	*oost410*	4,0	9,6	142	*71,4*
25	*corridor*	4,7	7,0	48	*16,7*
26	*oost330*	4,7	7,0	48	*16,7*
13	*corridor*	5,1	4,7	-8	*-6,7*
14	*oost410*	5,1	4,7	-8	*-6,7*
1	*corridor*	5,3	8,3	56	*18,7*
2	*oost450*	5,3	8,3	56	*18,7*
7	*corridor*	5,4	7,0	28	*5,4*
8	*oost430*	5,4	7,0	28	*5,4*
11	*corridor*	6,5	6,5	1	*-18,4*
12	*oost410*	6,5	6,5	1	*-18,4*
5	*corridor*	6,7	7,3	9	*-20,0*
6	*oost430*	6,7	7,3	9	*-20,0*
3	*corridor*	6,9	6,7	-3	*-5,5*
4	*oost450*	6,9	6,7	-3	*-5,5*
17	*corridor*	7,4	6,0	-19	*-16,3*
18	*oost370*	7,4	6,0	-19	*-16,3*
23	*corridor*	7,9	10,9	38	*32,7*
24	*oost330*	7,9	10,9	38	*32,7*
19	*corridor*	8,1	8,0	-2	*-0,2*
20	*oost350*	8,1	8,0	-2	*-0,2*
16	*oost370*	8,4	9,1	8	*-13,4*
15	*oost410*	8,6	9,1	5	*-13,0*
53	*corridor*	9,1	12,5	38	*17,1*
21	*corridor*	9,4	9,0	-5	*-0,5*
22	*oost350*	9,4	9,0	-5	*-0,5*
40	*oost370*	10,0	13,7	37	*2,6*
54	*corridor*	10,3	13,2	28	*-6,8*
58	*corridor*	11,0	20,9	90	*55,7*
51	*oost330*	11,1	11,8	6	*0,1*
56	*corridor*	12,0	13,5	13	*-18,0*
55	*corridor*	12,2	17,3	41	*9,1*
31	*oost450*	12,3	14,4	17	*4,2*
39	*oost410*	12,5	14,1	12	*5,2*
38	*oost410*	12,6	15,6	24	*11,1*
34	*oost430*	12,7	11,6	-9	*-20,1*
49	*oost330*	12,8	16,1	26	*7,6*
57	*corridor*	13,6	20,7	53	*14,5*
33	*oost430*	13,7	13,5	-2	*-7,1*
41	*oost370*	13,8	18,5	34	*21,5*
32	*oost450*	14,3	13,7	-4	*-5,1*
36	*oost430*	14,5	17,6	22	*-4,6*
37	*oost410*	15,1	17,6	17	*19,6*
35	*oost430*	15,5	18,1	16	*7,8*
44	*oost350*	15,7	15,1	-4	*4,0*
47	*oost330*	15,7	15,1	-4	*4,0*
52	*corridor*	15,8	27,3	73	*34,7*
50	*oost330*	16,2	15,5	-4	*-1,4*
30	*oost450*	16,2	10,6	-35	*-37,5*
46	*oost350*	16,4	21,2	30	*-4,3*
59	*corridor*	16,5	11,8	-28	*0,1*
27	*corridor*	17,0	22,7	33	*26,5*
45	*oost350*	17,3	18,5	7	*17,1*
48	*oost330*	17,8	17,2	-3	*-10,5*
43	*oost370*	20,3	22,8	12	*16,5*
29	*oost450*	21,2	11,2	-47	*-29,4*
60	*corridor*	21,5	16,2	-24	*0,2*
28	*corridor*	24,5	24,9	2	*-8,0*
42	*oost370*	24,8	25,4	2	*20,6*
				17,60 avg	***4,6*** avg

5 Nodes

Sample Pos	Zone	Non Opt	Opt	% Gain	*Metrics Gain Difference*
9	*oost430*	1,8	6,7	279	*42,3*
10	*oost410*	1,8	6,8	288	*73,7*
13	*corridor*	2,2	2,6	14	*-87,6*
14	*oost410*	2,2	3,8	70	*-31,7*
25	*corridor*	2,9	5,5	91	*71,0*
26	*oost330*	2,9	6,3	119	*98,4*
5	*corridor*	3,0	4,9	63	*57,7*
6	*oost430*	3,0	4,9	63	*57,7*
1	*corridor*	3,2	4,5	40	*11,3*
2	*oost450*	3,2	4,5	40	*11,3*
55	*corridor*	3,3	8,4	156	*33,4*
7	*corridor*	3,4	3,2	-4	*-56,8*
8	*oost430*	3,4	3,2	-4	*-56,8*
18	*oost370*	3,5	3,9	14	*-21,8*
20	*oost350*	3,5	3,8	10	*-41,1*
21	*corridor*	3,7	3,4	-10	*10,9*
22	*oost350*	3,7	3,4	-10	*10,9*
19	*corridor*	4,4	3,8	-13	*-32,4*
16	*oost370*	4,6	3,4	-27	*-11,3*
17	*corridor*	4,6	3,9	-14	*-16,5*
23	*corridor*	4,6	6,8	46	*69,8*
24	*oost330*	4,6	6,6	43	*66,8*
11	*corridor*	4,7	4,2	-10	*-1,1*
12	*oost410*	4,7	4,2	-10	*-1,1*
3	*corridor*	4,9	3,5	-27	*5,0*
4	*oost450*	4,9	3,5	-27	*5,0*
31	*oost450*	4,9	7,8	59	*41,4*
35	*oost430*	5,0	14,7	193	*117,1*
56	*corridor*	5,7	6,3	11	*-40,5*
53	*corridor*	5,9	5,9	1	*11,6*
36	*oost430*	6,0	11,4	89	*12,1*
15	*oost410*	6,0	2,6	-58	*-21,9*
40	*oost370*	6,0	7,9	31	*31,3*
32	*oost450*	6,3	8,2	29	*-17,7*
38	*oost410*	6,4	9,5	49	*1,4*
57	*corridor*	6,4	12,0	88	*56,9*
33	*oost430*	6,4	9,7	51	*53,8*
58	*corridor*	6,4	7,5	16	*40,0*
54	*corridor*	6,6	8,4	28	*3,4*
37	*oost410*	6,9	8,8	27	*-4,2*
34	*oost430*	7,1	8,8	23	*23,0*
47	*oost330*	7,3	4,3	-41	*-53,2*
59	*corridor*	7,3	5,3	-28	*-95,8*
39	*oost410*	7,4	7,9	7	*25,4*
42	*oost370*	7,8	9,4	21	*38,9*
41	*oost370*	8,1	7,0	-13	*-4,3*
45	*oost350*	8,3	7,0	-15	*7,6*
43	*oost370*	8,4	9,4	12	*32,5*
30	*oost450*	8,6	6,9	-20	*16,8*
49	*oost330*	8,9	5,5	-38	*-37,9*
51	*oost330*	8,9	13,2	48	*92,8*
60	*corridor*	9,5	3,0	-69	*-20,8*
48	*oost330*	9,7	3,0	-69	*-74,7*
52	*corridor*	9,7	6,3	-35	*-29,8*
50	*oost330*	10,3	6,5	-37	*-13,0*
46	*oost350*	10,4	8,1	-22	*-9,2*
44	*oost350*	10,5	7,4	-30	*-7,5*
27	*corridor*	10,9	4,7	-57	*-31,9*
29	*oost450*	11,8	8,6	-27	*3,1*
28	*corridor*	14,0	8,2	-42	*2,8*
				22,71 avg	***6,9*** avg

Figure 6.13: Distances to Nearest Localization Error

6.2.2. Assessing the kNN performance for k∈{3,5}

The above results help us estimate how the proposed optimizations may affect the localization performance, when 1 is used as the kappa coefficient in the kNN algorithm. Figure 6.14 presents a more direct evaluation of this performance, where now, 3 and 5 are used as the kappa coefficients. As in the previous case, this figure presents 3 tables (one for each setup-size) where both metrics (O1 denoting the first metric and O2 denoting the second one) are included. Each row corresponds to a specific evaluated cell position, the ID and true location of which can be found at the left columns. For every one of these cells, a location estimation has been performed in both the non-optimized and the two optimized scenarios, using the kNN algorithm and both kappa coefficients. The binary data within the table represent the binary state of correct or false localization. For example, considering the 5-nodes setup and the non-optimized case and after having found the 3 closest neighbors (i.e. k:3) of Cell 8 which belongs in reality to zone oost430, it was measured that most of these neighbors (i.e. at least 2) belonged to the same zone. Hence, a correct location estimation. However, in the optimized case using the minimum separation distance, the majority of the neighbors belonged to another zone and thus, in that case, the estimation was incorrect.

The summation of the correct estimations lies at the bottom of the columns, signifying better or worse overall localization performance. There, the green & red labels denote which coefficient produced the best performance. It is evident that, in every case, using as coefficient the value 3 (instead of 5) led to a better performance. This phenomenon, in accordance also with the previous results, may imply that using 1 as coefficient might lead to even better localization. Moreover, comparing the non-optimized to the optimized scenarios, for k=3, one can notice that the performance metrics led indeed to better localization performance. However, for k=5, the opposite was happening (i.e. the optimization was leading to worst localization). Moreover, the comparison between the two proposed metrics, reveals that the maximization of the minimum separation distance offered, in every case, better localization outcomes.

Finally, since a) the noise is considered in any case invariant and b) the localization is based on the kNN algorithm which basically measures distances (in RS units) to neighbors, it should be emphasized that any performance improvement mainly implies a better topology of this neighborhood, rather a pragmatic improvement in the localization accuracy. The latter could only be proved after considering also the noise which is a major influencing parameter.

		30 Nodes						15 Nodes						5 Nodes					
		k:3			k:5			k:3			k:5			k:3			k:5		
		N	O1	O2	N	O1	O2	N	O1	O2	N	O1	O2	N	O1	O2	N	O1	O2
1	corridor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	oost450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	oost450	1	0	0	1	0	0	0	1	1	1	0	0	1	0	0	1	0	0
5	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
6	oost430	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	1	0	0
7	corridor	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
8	oost430	1	1	1	1	0	0	1	1	1	0	0	1	1	0	1	1	0	0
9	oost430	1	0	1	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0
10	oost410	1	1	1	0	0	0	1	1	1	0	0	0	0	1	1	0	0	0
11	corridor	0	1	1	0	1	1	0	0	0	0	1	1	1	1	1	0	1	1
12	oost410	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0
13	corridor	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	1	0
14	oost410	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1
15	oost410	1	1	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	0
16	oost370	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0
17	corridor	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1	0
18	oost370	1	1	1	0	0	0	1	1	1	0	0	0	0	1	1	0	0	0
19	corridor	1	1	1	0	0	0	1	1	1	0	1	0	1	1	1	0	1	0
20	oost350	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
22	oost350	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	0	1	0
23	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
24	oost330	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1	0	1	0
25	corridor	0	1	1	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1
26	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1
27	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
29	oost450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
30	oost450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	oost450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
32	oost450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33	oost430	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
34	oost430	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
35	oost430	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	oost430	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
37	oost410	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38	oost410	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
39	oost410	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	oost370	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
41	oost370	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
42	oost370	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
43	oost370	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
44	oost350	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
45	oost350	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
46	oost350	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
47	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
48	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
49	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51	oost330	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
52	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1
53	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
54	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
55	corridor	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1
56	corridor	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1
57	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
58	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
59	corridor	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1
60	corridor	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		53	53	53	46	44	41	52	54	53	48	46	42	48	54	52	47	45	42

Figure 6.14: Localization based on more than 1 Neighbors

7

Conclusions

In this chapter, the knowledge that has been acquired throughout the graduation project shall be jointly considered for providing an educated answer to the formulated research questions. Furthermore, a discussion shall be made regarding the contribution of this thesis, our reflections and also our suggestions regarding future work.

7.1. Research Questions

1. How can the location distinctiveness be defined for an indoor positioning system?

Locations can be assessed using various metrics of uniqueness. Yet, since Location-enabled IPSs are typically based on radio signals, it can be inferred that their distinctiveness should also be assessed in terms of these.

Despite its polymorphous nature, a generalization can be made according to which, any location in 1, 2 or 3 dimensions can be considered as a superset of positions. Each one of these inherit as a thematic identifier, the single location to which they belong and is, at the same time, in a binding with a continuously varying Radio Signature which can be used to identify the position itself. However, due to the above variation, the identification process is not always accurate, leading often to positions that belong to other locations. As a result, the higher the probability this happens, the lower the location distinctiveness for the system is proved to be.

The margin within which these errors could occur, can be seen as an area of common radio signatures between the correct and the wrong position candidates. The smaller the area, the more distinct these radio signatures are, translating eventually to a better location distinctiveness. Therefore, the overall (from all possible positions) span where localization errors are not possible, defines how location distinctiveness can be perceived. According to this definition, it follows that

the location distinctiveness is a performance that does not describe individually a specific location, but instead, it is used to describe, in overall, a group of locations under specific radio conditions.

2. Which metric would be most suitable for measuring the location distinctiveness among different zone areas?

The location distinctiveness is bound to the way the radio signatures are utilized by the positioning techniques. Our proposal considers fingerprint-based approaches that respect the Euclidean space between the radio signatures. According to that, measuring the location distinctiveness requires that, first, a representative set of positions has been selected. This introduces the need of modeling the zones and converting the continuous space into a discrete one for extracting the positions that physically define the separation area between those zones. The same positions define also the separation borders of the Radio Signatures which shall be considered during the location distinctiveness measurement. To extract these positions in a meaningful way, an alpha shaping algorithm has to be utilized. Although computation geometry has various approaches to offer, a new simple and effective approach is proposed; namely, the Circular Manhattan Alpha Shaper.

To accurately measure the separation area of the Radio Signatures is highly impracticable. Therefore, alternative approaches should be utilized, introducing the corresponding margin of inaccuracy. A quite suitable solution is to perform a dense triangulation and thus, reduce the problem into calculating the sums of every generated triangle area. However, in the cases where more simplistic and efficient metrics are needed (e.g. for optimization purposes) two indirect metrics are proposed as the most suitable metrics. These do not measure separation areas, but instead, separation distances between positions that are on the border of a zone and within the separation area. For each one of these positions, this distance can be defined by the connection with the physically closest to it position that belongs to another zone.

- **Minimum Separation Distance:** This metric can identify the region where the Radio Separation is the smallest (or even the highest). That region is expected to have the lowest probability of offering correct localization estimations.
- **Product of the n Shortest Separation Distances:** Depending on the n coefficient, this metric can be quite similar to the previous one (even identical for n=1). For higher n values, this metric can offer bigger separation (in overall) distances.

3. Which radio propagation model would offer good accuracy-complexity ratio?

One of the most crucial parts of the optimization mechanism is the radio propagation modeling and that, because its accuracy is directly affecting the optimization's quality and performance. Therefore, empirical models that do not consider the physical geometry of the propagation environment should be avoided for such a task. Currently, the most accurate radio propagation models are considered to be the Finite-Difference Time-Domain ones. However, since these are computationally impossible to utilize, the next best candidates for our purpose are the ones which are based on Geometrical Optics. Both literature review and our results suggested that these can offer high levels of accuracy since they consider at an analytical level, both the electromagnetic properties and the propagation environments.

4. Which optimization algorithm should be utilized to support even large scale optimizations?

According to literature, both Simulated Annealing and Genetic Algorithms are powerful techniques that enable large scale optimizations. Since GAs have been favoured the most, this technique was eventually integrated into our optimization methodology. Although this renders comparisons between the two approaches impossible, yet, the performance of the implemented GA surpassed our expectations.

Two major difficulties were encountered during the development of this algorithm. Firstly, how to encode our specific problem into the philosophy of natural selection and secondly, how to balance its searching scope between random search and weighted search. Having done that, however, its speed performance proved to be quite remarkable. Even for large deployments, the solutions seem to converge sooner or later towards the same solution region which is impressive considering the vast search space.

Another property making GAs good candidates for large scale optimizations is their easy parallelizability. This enables the algorithm to scale according as much as the available computational resources allow.

5. How can the optimization results be evaluated?

To be aware of the extent of our optimization's performance is quite important and so, the optimization needs to be evaluated for each different node setup size. During this evaluation, the localization performance under some non-optimized deployment shall be compared with the localization performance under the corresponding optimized one. This performance can be computed for a set of sample cells and based on the kNN algorithm. The non-optimal deployments can reflect common practices that administrators often follow during IPS installations (i.e. regular placement). Yet, in an ideal scenario, our optimized solution should be compared to many different scenarios via a Monte Carlo process.

Having addressed the above sub-questions, we can now provide a concluding answer to the main research question:

To what extent can the placement of BLE nodes used for fingerprint-based positioning, be optimized to increase the location distinctiveness in an indoor environment?

To be able to measure the exact extent to which the placement of BLE nodes can be optimized and lead to better location distinctiveness, is quantitatively quite challenging. That, because location distinctiveness not only was approached from different directions (i.e. Separation Area, Separation Distances, etc), but also because in each case, the magnitude of the units was dependent via complicated relations from many other factors (e.g. cell-size, convergence of the GA, zone modeling, radio propagation modeling, etc).

Yet, the simulated evaluation showed that every optimized deployment introduced better location distinctiveness and thus, localization performance, when the estimations were based on the Nearest Neighbor. Moreover, it should be noted that a performance gain is also expected even when other localization techniques are used along with the fingerprinting (e.g. Bayesian Estimations). Finally, this optimization is expected to also have effect in the non-simulated cases. Otherwise, our radio propagation modelling was not accurate enough.

7.2. Reflection & Contribution

This graduation project introduces, for the first time, an optimization approach for Indoor Positioning Systems which are based on RSSI Fingerprinting and are orientated towards providing location estimations instead of exact positioning. Since localization refers to spatial information which lies in the heart of Geomatics, any improvement of these systems would be a contribution to the field of Geomatics. Yet, such systems are not solely relevant to the field of Geomatics, since the span of their applications is cross-disciplinary. Therefore, the contribution of our optimization is accordingly expansive. The evaluation of this approach showed that regardless the size of the node setup, our optimization does offer a localization improvement and thus, any installation of such systems should consider it.

Since our optimization requires a series of mechanisms (i.e. the radio simulation, the indoor modeling, the metric implementation and the search engine) which are quite expensive in terms of time & cost to develop or acquire, it is expected that our methodology may not be easily applicable. Yet, even without the development of an automated optimizer, our evaluation showed some deployment insights (e.g. placement of nodes across door openings) that are already useful to anyone being responsible for the installation of an IPS that is location enabled.

7.3. Future Work

The main objective of this graduation project was to develop a methodology for optimizing the deployment of BLE nodes, leading to better localization within an Indoor Positioning System. This objective has been reached, however, there are still closely relevant aspects worth researching further.

To begin with, in this project, the radio noise has not been considered. Therefore, modelling also the attenuation variations would introduce new paths and possibilities to our current work. It would not only become possible to research the minimum number of nodes required for a specific level of performance, but it would also enable the utilization of other localization techniques (e.g. Bayesian estimations).

Moreover, another research direction worth considering is to remove any placement constraints (e.g. only at walls) for the BLE nodes and to even enable the third dimension in case we have enough computational resources. In such case, highly accurate indoor models (e.g. based on point clouds) could also be used.

Finally, to put our methodology also to a practical test, we could also perform an evaluation within a real test scenario.

Bibliography

Adickes, Martin D. et al. (2002). "Optimization of indoor wireless communication network layouts". In: *IIE Transactions (Institute of Industrial Engineers)* 34.9, pp. 823–836. ISSN: 1573-9724. DOI: 10.1080/07408170208928915. URL: https://link.springer.com/article/10.1023/A:1015505023013.

Akl, Robert et al. (2006). "Indoor Propagation Modeling at 2.4 GHz for IEEE 802.11 Networks". In: *The Sixth IASTED International Multi-Conference on Wireless and Optical Communications*. Banff, AB, Canada. URL: http://www.cse.unt.edu/~rakl/ATL06.pdf.

Alsmady, Abdulsalam and Fahed Awad (2017). "Optimal Wi-Fi access point placement for RSSI-based indoor localization using genetic algorithm". In: *8th International Conference on Information and Communication Systems*. Irbid, Jordan: IEEE, pp. 287–291. ISBN: 978-1-5090-4243-2. DOI: 10.1109/IACS.2017.7921986. URL: https://ieeexplore.ieee.org/document/7921986.

Amzallag, David et al. (2005). "Cell planning of 4G cellular networks: algorithmic techniques and results". In: *2005 6th IEE International Conference on 3G and Beyond*. Washington, DC, USA: IET, pp. 501–505. ISBN: 1-4244-0816-4. URL: https://ieeexplore.ieee.org/abstract/document/4222833.

Baala, Oumaya et al. (2009). "The Impact of AP Placement in WLAN-Based Indoor Positioning System". In: *8th International Conference on Networks*. Gosier, Guadeloupe: IEEE, pp. 12–17. ISBN: 978-1-4244-3470-1. DOI: 10.1109/ICN.2009.50. URL: https://ieeexplore.ieee.org/document/4976643.

BuildingSmart (2016). *IFC Standard*. URL: http://www.buildingsmart-tech.org/specifications/ifc-overview.

Chang, Yang Lang et al. (2011). "A Parallel Simulated Annealing Approach to Band Selection for High-Dimensional Remote Sensing Images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.3, pp. 579–590. ISSN: 21511535. DOI: 10.1109/JSTARS.2011.2160048.

Chen, Guofeng et al. (2013). "Optimization of AP placement in indoor fingerprint positioning". In: *International Conference on ICT Convergence*. Jeju, South Korea: IEEE, pp. 98–100. ISBN: 978-1-4799-0698-7. DOI: 10.1109/ICTC.2013.6675316. URL: https://ieeexplore.ieee.org/document/6675316.

Chen, Qiuyun et al. (2014). "Placement of access points for indoor wireless coverage and fingerprint-based localization". In: *2013 IEEE International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*. Zhangjiajie,

China: IEEE, pp. 2253–2257. ISBN: 978-0-7695-5088-6. DOI: 10.1109/HPCC.and.EUC.2013.323. URL: https://ieeexplore.ieee.org/document/6832206.

Chen, Yingying et al. (2006). "A Practical Approach to Landmark Deployment for Indoor Localization". In: *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*. Reston, VA, USA: IEEE, pp. 365–373. ISBN: 1-4244-0626-9. DOI: 10.1109/SAHCN.2006.288441. URL: https://ieeexplore.ieee.org/document/4068139.

Dalla'Rosa, A. et al. (2011). "Deterministic Tool Based on Transmission Line Modelling and Kriging for Optimal Transmitter Location in Indoor Wireless Systems". In: *IET Microwaves, Antennas and Propagation* 5.13, pp. 1537–1545. ISSN: 17518725 17518733. DOI: 10.1049/iet-map.2010.0613. URL: https://ieeexplore.ieee.org/document/6086865.

Dalla'Rosa, Alexandre et al. (2008). "Optimal Indoor Transmitters Location Using TLM and Kriging Methods". In: *IEEE Transactions on Magnetics* 44.6, pp. 1354–1357. ISSN: 00189464. DOI: 10.1109/TMAG.2007.916239. URL: https://ieeexplore.ieee.org/document/4526837.

Diakité, Abdoulaye A. and Sisi Zlatanova (2018). "Spatial subdivision of complex indoor environments for 3D indoor navigation". In: *International Journal of Geographical Information Science* 32.2, pp. 213–235. ISSN: 13623087. DOI: 10.1080/13658816.2017.1376066. URL: https://doi.org/10.1080/13658816.2017.1376066.

Du, Xuan and Kun Yang (2017). "A Map-Assisted WIFI AP Placement Algorithm Enabling Mobile Device's Indoor Positioning". In: *IEEE Systems Journal* 11.3, pp. 1467–1475. ISSN: 19379234. DOI: 10.1109/JSYST.2016.2525814. URL: https://ieeexplore.ieee.org/document/7442075.

Eldeeb, Hassan et al. (2018). "Optimal placement of access points for indoor positioning using a genetic algorithm". In: *2017 12th International Conference on Computer Engineering and Systems (ICCES)*. Vol. 2018-Janua. Cairo, Egypt: IEEE, pp. 306–313. ISBN: 978-1-5386-1191-3. DOI: 10.1109/ICCES.2017.8275323. URL: https://ieeexplore.ieee.org/document/8275323.

ESRI (1998). *ESRI Shapefile Technical Description*. Tech. rep. URL: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

Fang, Shih Hau and Tsung Nan Lin (2010). "A novel access point placement approach for WLAN-based location systems". In: *IEEE Wireless Communications and Networking Conference, WCNC*. IEEE. ISBN: 978-1-4244-6398-5. DOI: 10.1109/WCNC.2010.5506586. URL: https://ieeexplore.ieee.org/document/5506586.

Faragher, Ramsey and Robert Harle (2015). "Location Fingerprinting With Bluetooth Low Energy Beacons". In: *IEEE Journal on Selected Areas in Communications* 33.11, pp. 2418–2428. ISSN: 1558-0008. DOI: 10.1109/JSAC.2015.2430281. URL: https://ieeexplore.ieee.org/document/7103024.

Farkas, Károly et al. (2013). "Optimization of Wi-Fi Access Point Placement for Indoor Localization". In: *Informatics & IT Today* 1.1, pp. 28–33. ISSN: 1339-147X.

Ficco, Massimo et al. (2013). "Calibrating Indoor Positioning Systems with Low Efforts". In: *IEEE Transactions on Mobile Computing* 13.4, pp. 737–751. DOI: 10.1109/TMC.2013.29. URL: https://ieeexplore.ieee.org/document/6473800.

Fortune, Steven J. et al. (1995). "WISE design of indoor wireless systems: practical computation and optimization". In: *IEEE Computational Science and Engineering* 2.1, pp. 58–68. DOI: 10.1109/99.372944. URL: https://ieeexplore.ieee.org/document/372944.

Gomez, Carles et al. (2012). "Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology". In: *Sensors (Switzerland)* 12.9, pp. 11734–11753. ISSN: 14248220. DOI: 10.3390/s120911734. URL: http://www.mdpi.com/1424-8220/12/9/11734/pdf.

Grubisic, S. et al. (2009). "Optimization Model for Antenna Positioning in Indoor Environments Using 2-D Ray-Tracing Technique Associated to a Real-Coded Genetic Algorithm". In: *IEEE Transactions on Magnetics* 45.3, pp. 1626–1629. ISSN: 1941-0069. DOI: 10.1109/TMAG.2009.2012760. URL: https://ieeexplore.ieee.org/document/4787442.

Haagmans, G.G. (2017). "A statistical analysis on the system performance of a Bluetooth Low Energy indoor positioning system in a 3D environment". PhD thesis. Delft: Delft University of Technology. URL: https://repository.tudelft.nl/islandora/object/uuid%3Af7646f3b-9a99-4207-97a6-20ef83c6a571.

Hara, Shinsuke and Takahiro Fukumura (2008). "Determination of the placement of anchor nodes satisfying a required localization accuracy". In: *2008 IEEE International Symposium on Wireless Communication Systems*. Reykjavik, Iceland: IEEE, pp. 128–132. ISBN: 978-1-4244-2488-7. DOI: 10.1109/ISWCS.2008.4726032. URL: https://ieeexplore.ieee.org/document/4726032.

He, Ying et al. (2011). "Rapid Deployment of APs in WLAN Indoor Positioning System". In: *6th International ICST Conference on Communications and Networking in China (CHINACOM)*. Harbin, China: IEEE, pp. 268–273. ISBN: 978-1-4577-0101-6. DOI: 10.1109/ChinaCom.2011.6158161. URL: https://ieeexplore.ieee.org/document/6158161.

Huszák, Árpád et al. (2012). "Investigation of WLAN Access Point Placement for Indoor Positioning". In: *IFIP International Federation for Information Processing*. Springer, Berlin, Heidelberg, pp. 350–361. ISBN: 978-3-642-32808-4. DOI: 10.1007/978-3-642-32808-4{_}32. URL: https://link.springer.com/chapter/10.1007/978-3-642-32808-4_32.

International Telecommunication Union (2016). *Radio Regulations*. Tech. rep. URL: http://handle.itu.int/11.1002/pub/80da2b36-en.

Ji, Zhong et al. (2002). "Methods for Optimizing the Location of Base Stations for Indoor Wireless Communications". In: *IEEE Transactions on Antennas and Prop-*

agation 50.10, pp. 1481–1483. DOI: 10.1109/TAP.2002.802155. URL: https://ieeexplore.ieee.org/document/1137545.

Kang, Sung-lien et al. (2013). "Wireless LAN Access Point Location Planning". In: *Proceedings of the Institute of Industrial Engineers Asian Conference*. Springer, Singapore, pp. 907–914. DOI: 10.1007/978-981-4451-98-7{_}108. URL: https://link.springer.com/chapter/10.1007/978-981-4451-98-7_108.

Karl, Holger and Andreas Willig (2005). *Protocols and Architectures for Wireless Sensor Networks*, p. 526. ISBN: 0-470-09510-5. URL: https://ieeexplore.ieee.org/servlet/opac?bknumber=8040281.

Kolodziej, Krzysztof W. and Johan. Hjelm (2006). *Local Positioning Systems - LBS Applications and Services*. Taylor & Francis Group, pp. 120–122. ISBN: 978-0-8493-3349-1.

Kondee, Kittipob et al. (2015). "A Novel technique for Reference Node Placement in Wireless Indoor Positioning Systems based on Fingerprint Technique". In: *Computer and Information Technology* 9.2, pp. 131–141. URL: https://www.tci-thaijo.org/index.php/ecticit/article/view/54416.

Kouhbor, S. et al. (2006). "Coverage in WLAN with Minimum Number of Access Points". In: *63rd Vehicular Technology Conference*. Vol. 3. c. Melbourne, Vic., Australia: IEEE, pp. 1166–1170. ISBN: 0-7803-9391-0. DOI: 10.1109/VETECS.2006.1683018. URL: https://ieeexplore.ieee.org/document/1683018.

Laitinen, Elina and Elena-Simona Lohan (2016). *Access Point topology evaluation and optimization based on Cramér-Rao Lower Bound for WLAN indoor positioning*. Tech. rep. Barcelona, Spain. DOI: 10.1109/ICL-GNSS.2016.7533850. URL: https://ieeexplore.ieee.org/document/7533850.

Ledesma, Sergio et al. (2012). *Simulated Annealing Evolution*. ISBN: 978-953-51-0710-1. DOI: 10.5772/50176. URL: https://www.intechopen.com/books/simulated-annealing-advances-applications-and-hybridizations/simulated-annealing-evolution.

Lee, Jin-Shyan et al. (2007). "A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi". In: *IECON 2007 - 33rd Annual Conference of the IEEE Industrial Electronics Society*. Taipei, Taiwan: IEEE, pp. 46–51. ISBN: 1-4244-0783-4. DOI: 10.1109/IECON.2007.4460126. URL: https://ieeexplore.ieee.org/document/4460126.

Lee, Jong-Hyouk et al. (2007). "An Optimal Access Points Allocation Scheme Based on Genetic Algorithm". In: *Future Generation Communication and Networking (FGCN 2007)*. Jeju, South Korea: IEEE. ISBN: 0-7695-3048-6. DOI: 10.1109/FGCN.2007.66. URL: https://ieeexplore.ieee.org/document/4426203.

Li, Dong et al. (2015). "Measurement-based access point deployment mechanism for indoor localization". In: *Global Communications Conference*. San Diego, CA, USA: IEEE, pp. 1–6. ISBN: 978-1-4799-5952-5. DOI: 10.1109/GLOCOM.2015.7417107. URL: https://ieeexplore.ieee.org/document/7417107.

Li, Ki-Joune et al. (2019). *OGC IndoorGML: A Standard Approach for Indoor Maps*. Elsevier Inc., pp. 187–207. ISBN: 9780128131893. DOI: 10.1016/B978-0-12-813189-3.00010-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780128131893000101.

Liang, Hui et al. (2012). "A Novel Transmitter Placement Scheme Based on Hierarchical Simplex Search for Indoor Wireless Coverage Optimization". In: *IEEE Transactions on Antennas and Propagation* 60.8, pp. 3921–3932. ISSN: 1558-2221. DOI: 10.1109/TAP.2012.2201081. URL: https://ieeexplore.ieee.org/abstract/document/6204053.

Liao, Lin et al. (2011). "Two birds with one stone: Wireless access point deployment for both coverage and localization". In: *IEEE Transactions on Vehicular Technology* 60.5, pp. 2239–2252. ISSN: 00189545. DOI: 10.1109/TVT.2011.2109405. URL: https://ieeexplore.ieee.org/document/5710001.

Liu, Hui et al. (2007). "Survey of wireless indoor positioning techniques and systems". In: *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 37.6, pp. 1067–1080. ISSN: 1558-2442. DOI: 10.1109/TSMCC.2007.905750. URL: https://ieeexplore.ieee.org/abstract/document/4343996.

Liu, Liu et al. (2019). "Indoor Routing on Logical Network Using Space Semantics". In: *ISPRS International Journal of Geo-Information* 8.3, p. 126. ISSN: 2220-9964. DOI: 10.3390/ijgi8030126. URL: https://www.mdpi.com/2220-9964/8/3/126.

Luo, Meiling (2013). "Indoor radio propagation modeling for system performance prediction". PhD thesis. URL: http://theses.insa-lyon.fr/publication/2013ISAL0074/these.pdf.

Maksuriwong, K. et al. (2003). "Wireless LAN access point placement using a multi-objective genetic algorithm". In: *International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance*. Washington, DC, USA: IEEE, pp. 1944–1949. ISBN: 0-7803-7952-7. DOI: 10.1109/ICSMC.2003.1244696. URL: https://ieeexplore.ieee.org/document/1244696.

Mc Gibney, Alan et al. (2010). "A wireless local area network modeling tool for scalable indoor access point placement optimization". In: *Proceedings of the 2010 Spring Simulation Multiconference on*. San Diego, CA, USA: Society for Computer Simulation International. ISBN: 978-1-4503-0069-8. DOI: 10.1145/1878537.1878707. URL: http://portal.acm.org/citation.cfm?doid=1878537.1878707.

McKay, J.B. and M. Pachter (1997). "Geometry optimization for GPS navigation". In: *Proceedings of the 36th IEEE Conference on Decision and Control*. Vol. 5. December. San Diego, CA, USA: IEEE, pp. 4695–4699. ISBN: 0-7803-4187-2. DOI: 10.1109/CDC.1997.649738. URL: https://ieeexplore.ieee.org/document/649738.

Meng, Weixiao et al. (2012). "Optimized Access Points Deployment for WLAN Indoor Positioning System". In: *Wireless Communications and Networking Conference*. Shanghai, China: IEEE, pp. 2457–2461. ISBN: 978-1-4673-0437-5. DOI: 10.1109/WCNC.2012.6214209. URL: https://ieeexplore.ieee.org/document/6214209.

Minkara, Rania and Peter Shepherd (2014). "Optimising the location of wireless base stations within a dynamic indoor environment". In: *The 8th European Conference on Antennas and Propagation*. The Hague, Netherlands: IEEE, pp. 2101–2105. ISBN: 978-8-8907-0184-9. DOI: 10.1109/EuCAP.2014.6902222. URL: https://ieeexplore.ieee.org/document/6902222.

Moreno, Jose et al. (2015). "Design of Indoor WLANs: Combination of a ray-tracing tool with the BPSO method." In: *IEEE Antennas and Propagation Magazine* 57.6, pp. 22–33. ISSN: 10459243. DOI: 10.1109/MAP.2015.2480078. URL: https://ieeexplore.ieee.org/document/7362120.

Nagy, Lajos and Lóránt Farkas (2000). "Indoor base station location optimization using genetic algorithms". In: *11th International Symposium on Personal, Indoor and Mobile Radio Communication*. IEEE, pp. 843–846. ISBN: 0-7803-6463-5. DOI: 10.1109/PIMRC.2000.881541. URL: https://ieeexplore.ieee.org/document/881541.

Nykamp, DQ (2019). *Calculating the area under a curve using Riemann sums*. URL: https://mathinsight.org/calculating_area_under_curve_riemann_sums.

OGC (2012). *OGC City Geography Markup Language (CityGML) Encoding Standard*. URL: http://www.opengeospatial.org/standards/citygml.

— (2015). *Ogc Kml 2.3*. URL: http://docs.opengeospatial.org/is/12-007r2/12-007r2.html.

— (2016). *OGC IndoorGML Corrigendum*. URL: http://www.opengeospatial.org/standards/indoorgml.

Politi, Alexandre A. et al. (2016). "An optimization model for the indoor access point placement problem with different types of obstacles". In: *2015 Latin America Congress on Computational Intelligence (LA-CCI)*. Curitiba, Brazil: IEEE. ISBN: 978-1-4673-8418-6. DOI: 10.1109/LA-CCI.2015.7435979. URL: https://ieeexplore.ieee.org/document/7435979.

Redondi, Alessandro E.C. and Edoardo Amaldi (2013). "Optimizing the placement of anchor nodes in RSS-based indoor localization systems". In: *12th Annual Mediterranean Ad Hoc Networking Workshop*. Ajaccio, France: IEEE, pp. 8–13. ISBN: 978-1-4799-1004-5. DOI: 10.1109/MedHocNet.2013.6767403. URL: https://ieeexplore.ieee.org/abstract/document/6767403.

Remley, Kate A. et al. (2000). "Improving the accuracy of ray-tracing techniques for indoor propagation modeling". In: *IEEE Transactions on Vehicular Technology* 49.6, pp. 2350–2358. ISSN: 1939-9359. DOI: 10.1109/25.901903. URL: https://ieeexplore.ieee.org/document/901903.

Rengarajan, Balaji and Gustavo de Veciana (2005). *Optimizing Placement of Wireless Access Points*. Tech. rep. Austin, Texas: The University of Texas at Austin.

Roberto, Battiti et al. (2003). *Optimal wireless access point placement for location-dependent services*. Tech. rep., pp. 1–12. DOI: 10.1.1.10.6631.

Saleh, Hussain Aziz and Rachid Chelouah (2004). "The design of the global navigation satellite system surveying networks using genetic algorithms". In: *Engineering Applications of Artificial Intelligence* 17.1, pp. 111–122. ISSN: 09521976. DOI: 10.1016/j.engappai.2003.11.001.

Sharma, Chhavi et al. (2010). "Access point placement for fingerprint-based localization". In: *12th IEEE International Conference on Communication Systems*. Singapore, Singapore: IEEE, pp. 238–243. ISBN: 978-1-4244-7006-8. DOI: 10.1109/ICCS.2010.5686092. URL: https://ieeexplore.ieee.org/document/5686092.

Staats, Bart (2017). "Identification of walkable space in a voxel model, derived from a point cloud and its corresponding trajectory". PhD thesis. TU Delft. URL: http://resolver.tudelft.nl/uuid:6a827a88-ce09-43f1-adb9-da886448a1fc.

Talau, Marcos et al. (2013). "Solving the base station placement problem by means of swarm intelligence". In: *2013 IEEE Symposium on Computational Intelligence for Communication Systems and Networks (CIComms)*, pp. 39–44. ISBN: 978-1-4673-5903-0. DOI: 10.1109/CICommS.2013.6582852. URL: https://ieeexplore.ieee.org/document/6582852.

Taufiq, Muhammad et al. (2011). "Wireless LAN access point placement based on user mobility". In: *Wireless Personal Communications* 60.3, pp. 431–440. ISSN: 1572-834X. DOI: 10.1007/s11277-011-0300-0. URL: https://link.springer.com/article/10.1007/s11277-011-0300-0.

Vilović, Ivan and Nikša Burum (2014). "Location Optimization of WLAN Access Points Based on a Neural Network Model and Evolutionary Algorithms". In: *Automatika – Journal for Control, Measurement, Electronics, Computing and Communications* 55.3, pp. 317–329. ISSN: 1848-3380. DOI: 10.7305/automatika.2014.12.556. URL: https://automatika.korema.hr/index.php/automatika/article/view/556.

Voronov, Roman Vladimirovich (2017). "The problem of optimal placement of access points for the indoor positioning system". In: 13.1, pp. 61–73. DOI: 10.21638/11701/spbu10.2017.106.

Wertz, P et al. (2004). "Automatic optimization algorithms for the planning of wireless local area networks". In: *60th Vehicular Technology Conference*. Los Angeles, CA, USA: IEEE, pp. 3010–3014. ISBN: 0-7803-8521-7. DOI: 10.1109/VETECF.2004.1400613. URL: https://ieeexplore.ieee.org/document/1400613.

Woolley, Martin (2019). *Bluetooth Direction Finding - A Technical Overview*. Tech. rep. 1.0. Bluetooth SIG, Inc. URL: https://www.bluetooth.com/bluetooth-resources/bluetooth-direction-finding/.

Worboys, Michael (2011). "Modeling indoor space". In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness - ISA '11*, pp. 1–6. DOI: 10.1145/2077357.2077358. URL: http://dl.acm.org/citation.cfm?doid=2077357.2077358.

Xenakis, D., M. Meijers, et al. (2019). "Placement optimization of positioning nodes: Maximizing the distinction of indoor zones". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42.2/W13, pp. 909–914. ISSN: 16821750. DOI: 10.5194/isprs-archives-XLII-2-W13-909-2019. URL: https://doi.org/10.5194/isprs-archives-XLII-2-W13-909-2019.

Xenakis, D. and E. Verbree (2018). *Translating BLE Received Signals across different Smartphones*. Tech. rep. Delft: TU Delft. URL: http://resolver.tudelft.nl/uuid:e5768343-3143-49e0-b6eb-3239e064c96d.

Yoon, Yourim and Yong Hyuk Kim (2013). "An Efficient Genetic Algorithm for Maximum Coverage Deployment in Wireless Sensor Networks". In: *IEEE Transactions on Cybernetics* 43.5, pp. 1473–1483. ISSN: 21682267. DOI: 10.1109/TCYB.2013.2250955. URL: https://ieeexplore.ieee.org/document/6497561.

Yun, Zhengqing et al. (2008). "An integrated method of ray tracing and genetic algorithm for optimizing coverage in indoor wireless networks". In: *IEEE Antennas and Wireless Propagation Letters* 7, pp. 145–148. ISSN: 15361225. DOI: 10.1109/LAWP.2008.919358. URL: https://ieeexplore.ieee.org/document/4456057.

Zhang, Rui et al. (2014). "The optimal placement method of anchor nodes toward RSS-based localization systems". In: *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*. Hefei, China: IEEE. ISBN: 978-1-4799-7339-2. DOI: 10.1109/WCSP.2014.6992059. URL: https://ieeexplore.ieee.org/document/6992059.

Zlatanova, S. et al. (2014). *Space subdivision for indoor applications*. January. ISBN: 9789077029374. DOI: 10.13140/2.1.2914.2081.