OPTIMIZATION OF BINARY FINGERPRINTING DESIGN FOR POSITIONING SYSTEMS.

OPTIMIZACIÓN DEL DISEÑO DE ‘FINGERPRINTING’ BINARIO PARA SISTEMAS DE LOCALIZACIÓN.

Trabajo de Fin de Grado de Ingeniería de Tecnologías y Servicios de Telecomunicación

Universidad Politécnica de Madrid

Curso 2017-2018

Autor: Miguel Omeñaca Muro
Tutor: Luca Reggiani

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

Fecha y lugar de lectura: Milán, 17 de Julio de 2018
Resumen

El trabajo de fin de grado propuesto está relacionado con técnicas de localización, concretamente con las que utilizan ”Fingerprinting” (FP). Este proyecto se basa en el diseño propuesto por M. Mizmizi y L. Reggiani en dos publicaciones recientes (véase [1] y [2]) con el fin de analizar posibles variaciones y mejoras. Los autores proponen reducir el número de niveles de cuantificación del indicador de intensidad de la señal en recepción (RSSI) hasta su representación binaria con el fin de limitar la complejidad de los sistemas que utilizan FP equipados con balizas con tecnologías para el Internet de las Cosas (IoT).

El trabajo realizado está enmarcado en una línea de investigación cuyo objetivo principal es el de diseñar sistemas de localización de bajo coste y de baja complejidad en el contexto de comunicaciones ”machine to machine” (M2M) y con un gran número de dispositivos en escenarios de difícil propagación (espacios interiores, almacenes, fábricas...). Los objetivos específicos del proyecto son los siguientes:

- Mejora del diseño de la técnica de ”Fingerprinting” binaria utilizando un novedoso proceso de cuantificación.
- Diseño del algoritmo extendido a medidas de señal alternativas.
- Validación experimental de los resultados.

El proyecto ha sido realizado durante los meses de Marzo (M0) a Julio (M4) de 2018 con la siguiente planificación y metodología:

- M0-M1: Búsqueda bibliográfica y estudio de trabajos relacionados. Preparación de un análisis, incluido en esta memoria, de los sistemas de ”Fingerprinting” más modernos.
- M2: Diseño de la técnica de posicionamiento y preparación del simulador.
- M3: Actividad experimental y simulaciones.
- M4: Análisis de los resultados y fin de la memoria.
A modo de orientación, señalar que hemos realizado reuniones cada dos semanas y cuando ha resultado ser necesario en base al progreso del trabajo.

Las fases del proyecto han seguido el esquema anterior de planificación. Es decir, primero se ha realizado una fase de documentación sobre las técnicas de posicionamiento en general y después sobre las que únicamente corresponden a “Fingerprinting” (FP). Al cabo de ésta, se ha comenzado a redactar el primer capítulo de esta memoria (véase 1). Después, se ha llevado a cabo una fase introductoria al simulador para conocer su funcionamiento, sus parámetros y detalles. Para ello, se han realizado simulaciones del sistema original de FP binario variando ciertos parámetros y analizando su influencia en el rendimiento del sistema. Todo esto ha quedado plasmado en el segundo capítulo (véase 2) de esta memoria. La tercera fase del proyecto ha consistido en una fase de desarrollo y de diseño. Se han propuesto unas mejoras respecto al sistema original binario y se han realizado las modificaciones oportunas en el simulador. Además, se han simulado varios escenarios para obtener distintos resultados a analizar. La cuarta fase ha sido una fase de análisis de los resultados. Tanto esta fase como la anterior, han quedado retratadas en el tercer capítulo (véase 3). Por último, la fase final ha consistido en la elaboración de este documento tras la obtención de las conclusiones oportunas.

En resumen podemos concluir que los resultados de nuestro trabajo han sido satisfactorios ya que el objetivo principal, es decir, optimizar la técnica de FP binaria propuesta, se ha realizado correctamente. Para ello, se ha utilizado un novedoso proceso de cuantificación ternario adaptado al sistema anterior. Adicionalmente se ha extendido y adaptado el algoritmo de localización a esta nueva representación y además se ha mejorado dicho algoritmo. La combinación de ambos aportes ha resultado mejorar la precisión del sistema.

PALABRAS CLAVE: Fingerprinting, Sistemas de localización, Representación Binaria, Cuantificación, Representación Ternaria, Estimación de la posición, Redes de sensores, Bluetooth Low Energy, Balizas.
Abstract

Location Based Services (LBS) have become an essential part of our day-to-day life. One clear example is the well-known Global Positioning System (GPS) developed around 1970 and which has turned out to be a successful solution for outdoor environments. In fact, it is still used by the vast majority of the devices around the world. More recently, the scientific community has focused on Indoor Positioning Systems (IPS) which still lack of global solutions.

A great variety of IPS have been developed using different technologies like Ultra-Sounds, Infra-red, Radio Frequency Waves and more, as well as several positioning techniques. However still none of these systems completely meet the requirements of the market. Schemes having their own dedicated sensor network (WSN) and using a low-cost technology like Bluetooth do stand out from the rest. That is because they achieve an interesting trade-off between performance and cost.

The parameter mostly used for location purposes is the Received Signal Strength Indicator (RSSI) for its simplicity and availability in each device; other measures such as the Angle of Arrival (AOA) or Time of Arrival are more complicated and they usually require additional or dedicated hardware. It is quite often to make advantage of the RSSI parameter by using the so-called Fingerprinting technique because it enables to achieve the best performance.

Fingerprinting is now a widely-used technique for IPS that has been further developed by researchers in order to improve its performance. This Bachelor Thesis proposes an optimization of a specific system using this technique in combination with a Binary quantization scheme. The proposed new solutions are the following ones:

- a positioning estimation technique that improves the one used in the original system;
- the adoption of a Ternary quantization labelling method enhancing the overall performance of the scheme;
The combination of these solutions can be exploited to obtain a small average error on the position estimate.

Finally, a simulator has been adapted from the previous one that used the Binary technique. It has allowed to validate the proposed new model and to discover new exploitable possible optimizations.

**KEY WORDS:** Fingerprinting, Indoor Positioning Systems, Binary representation, Quantization, Ternary representation, Position Estimation, Sensor Networks, Bluetooth Low Energy, Beacons.
# Contents

Introduction .......................................................... 2

1 The state of the art in fingerprinting-based indoor positioning systems .... 4
  1.1 Criteria for evaluating indoor positioning systems ..................... 5
  1.2 Introduction to fingerprinting techniques ............................ 6
  1.3 Reference point clustering and coarse localization ................... 7
  1.4 Radio map construction .............................................. 9
  1.5 Density and weight estimation ...................................... 12
  1.6 Additional recent approaches ...................................... 13

2 Binary fingerprinting-based indoor positioning system .................. 20
  2.1 Binary fingerprinting technique .................................. 21
  2.2 Channel model ..................................................... 26
  2.3 Real scenario simulations and results ............................. 27

3 Optimization of the binary fingerprinting-based indoor positioning system 31
  3.1 Introduction to the optimized system ............................... 32
  3.2 Modifications and positioning algorithm .......................... 34
  3.3 Numerical results and comparisons ................................. 35

Conclusions and future work .......................................... 40

Ringraziamenti ......................................................... 42

Bibliography .......................................................... 43
Introduction

In recent years, the scientific community has been interested in the design and development of localization systems that can operate in indoor environments and at the same time can achieve a degree of precision, reliability and cost comparable to the well-known Global Positioning System (GPS). These systems, often known as Indoor Positioning Systems (IPS), will allow to advance in fields such as location-aware, pervasive computing, ambient intelligence and home automation. In addition, they will encourage the deployment of location-based services (LBS).

The spread of IPS will be heavily determined by their trade-off between cost and performance. Furthermore, these systems should adopt low-cost and low-power technologies implying low maintenance and requiring minimum new infrastructure with respect to those already present and installed for Wireless Personal or Local Area Networks (WPAN / WLAN).

There are many available technologies matching the above-mentioned necessities but that still need to achieve high-accuracy in indoor environments. For instance it is the case of Ultra Wide Band (UWB) ranging and Radio Frequency Identification (RFID) systems which additionally would need new infrastructure and radios into mobile devices.

This Bachelor Thesis is related to IPS and specially to those based on a sensor network or WPAN. The technology adopted is the Bluetooth Low Energy (BLE) since it has many advantages such as the low power consumption and the large-scale availability. The work focuses on a sensor network-based IPS employing a novel Fingerprinting technique that uses Binary quantization for labelling the fingerprints.
This work is organized as follow. Chapter 1 presents the state of art in fingerprinting-based IPS and reviews some of the most recent techniques and approaches adopted in these systems. In Chapter 2, the Binary-fingerprinting based IPS is thoroughly described and analysed: it is the foundation of this project. Chapter 3 represents the core of this Bachelor thesis: it describes the optimization made to the Binary-fingerprinting technique and its results. In this chapter some simulations have been carried out in order to confirm the improvements and benefits of the new modifications.
Chapter 1

The state of the art in fingerprinting-based indoor positioning systems

Over the last years, fingerprinting has been considered as the most popular technique for indoor location estimation. Originally it was conceived by Microsoft Researchers while developing the Radar System. Nowadays it is widely used and the scientific community keeps developing new approaches to face Indoor Positioning Systems (IPS) related issues.

This chapter gives a general idea of the most recent fingerprinting-based IPS which can either be commercial or research-oriented solutions. These solutions have different mechanisms and generally focus on specific features that need to be developed. That is why, beforehand, it is required to determine the criteria used for the comparisons and evaluations of the systems.

From now on, following the criteria mentioned before there will be a classification, comparison and evaluation of the different positioning systems presented throughout this chapter. Note that the majority of the solutions are taken from two different articles on recent advances (i.e. [3] and [4]) and then further developed.
1.1 Criteria for evaluating indoor positioning systems

Different criteria is required to evaluate Indoor Positioning Systems (IPS). The following ones are the most essential and generic when dealing with the deployment of an IPS:

1. **Accuracy**: Accuracy in the context of IPS is also known as location error. It is the most important feature when it comes to the analysis and evaluation of such systems. Generally the mean distance error which is the average Euclidean distance between the estimated location and the true location is employed as the performance metric. It is clear that the higher the accuracy, the better the system but usually a tradeoff between accuracy and all the others characteristics is needed in order to perform well.

2. **Complexity**: Complexity has two main interesting aspects. The first one is the computing difficulty of the positioning algorithm. If it is performed in a server then there is no problem but in the general case it is performed in a mobile device which brings problems related to the lack of processing power and battery life. As a consequence for IPS, it is preferable to use low calculation complexity. The other aspect is associated to the human resources and efforts needed during the deployment and maintenance of the IPS. In general IPS have a fast set-up requiring a low number of fixed infrastructure components.

3. **Latency**: Latency in an IPS is key to the overall behaviour of the system. It will interfere during the measurement phase and might slow down all the positioning process.

4. **Cost**: Like in most systems cost is a feature to bear in mind. It depends on multiple factors like money, time and space related to the system installation and maintenance, mobile devices and so on.

The criteria mentioned above will be used to evaluate and compare IPS in the coming sections.
1.2 Introduction to fingerprinting techniques

Fingerprinting techniques are very common because they are reliable and simple. In addition, there are several measurements that can be used in such systems. Usually, the received signal strength indicator (RSSI) is the choice but signal to noise ratio (SNR), channel impulse response, angle of arrival (AOA), time of arrival (TOA) can also be exploited. The fingerprinting method is always divided in two different phases: the off-line phase and the on-line phase. In the following, both phases will be explained based on one specific paper (i.e. [5]) as well as on a thesis (i.e. [6]).

- **Off-line phase**: it is known as the calibration stage and basically it is in charge of the collection and storage of all fingerprints (FPs) in the database. A test mobile station (MS) moves in a grid with a large set of positions $x_j$ covering the whole indoor environment of interest recording levels from different base stations (BSs), $\{p_i(x_j)\}$, with $\{i = 1, ..., n\}$.

- **On-line phase**: it is known as the localization stage and this is the step in which the system localizes the target. The system reads a set of new signals $\{r_i\}$ and compares it with the previously recorded set. The position that matches the best is chosen as the location estimate. For this matter many techniques can be used such as machine learning related ones. The most common choice to estimate the location is the Nearest Neighbor (NN) algorithm which simply picks the grid point $x_j$ minimizing the distance (usually Euclidean) between the measured RSSI vector and the FPs in the database.

$$\hat{x} = \text{argmin}\{z_j^2\} \text{ with } z_j^2 = \sum_{i=1}^{n}(r_i - p_i(x_j))^2$$

Another very common estimation technique is the Weighted K-Nearest Neighbor (WKNN) one. Essentially it is the same idea than in the NN algorithm but including weights to calibrate the points. A simple weight can be the inverse of the RSSI norms. Besides, it is important to bear in mind that fingerprinting techniques do not distinguish between line of sight (LoS) and non-line-of-sight (NLoS) and therefore are robust to these conditions.

In the following sections, different fingerprinting techniques classified in groups and categories will be presented.
1.3 Reference point clustering and coarse localization

It is well known that radio signal strength (RSS) fingerprints (FPs) depend on the environment in which they are collected and on the number of available access points (APs). Recent works have focused on limiting the positioning algorithm to a subset of reference points (RPs) in order to limit the search space and strengthen the search in the specific subset of RPs. This is known as radio map clustering or spatial filtering. It should be mentioned that it is an off-line process grouping certain RPs following a specific criteria, generally on similar metric. Basically the main idea is to make a clustering and coarse localization stage allowing to feed a fine localization stage. In this section multiple solutions will be introduced and briefly discussed.

1. **Use of binary AP coverage:** This spatial filtering method assumes that neighbouring RPs receive similar FPs. The approach taken here really emphasizes in reliability. That is to say, an AP is reliable for a given RP if the FPs are above a specific threshold. Additionally, if an AP is reliable its coverage indicator will be 1, otherwise it will be 0. Then, reliable APs for a given RP are stored in a binary coverage vector and afterwards in a coverage vector defined from on-line measurements. The coarse localization is performed by taking a subset of RPs following a certain condition usually related to the Hamming distance of the two vectors.

2. **K-means clustering:** The idea here is to find clusters and their centroids iteratively. Then, these centroids are updated at each iteration. Besides, the number of clusters is defined through a training set as follows. First K RPs are selected as centroids for each cluster. Afterwards when all the RPs are assigned the centroid of each cluster is recalculated by taking the average of the signal strength of the RPs belonging to it. There is a late evolution called Block-based weighted clustering which uses an objective function to compute the weight for each cluster. The objective function has the task to find the optimal set of centroids by minimizing the least square distance between the RP and centroids.
3. **Affinity propagation**: This approach takes as input a set of real-valued similarities used to determine the convenience of a RP as a centroid. For instance when the focus is on minimising the square error, the similarity is set as the negative square error. This method is based on iterative messages between nodes and centroids. These messages integrate a competition of centroids for the RPs in an iterative manner that converges to a final decision defining clusters and centroids. As it is explained above, this approach is quite similar to the K-means one in the sense that clusters are selected for the coarse localization and later used for fine localization. Nevertheless it is also different because it has no initialization.

4. **Splitting-based clustering**: This method is different from the other ones already mentioned because it is not based in similarities. The process starts by taking the whole area and splits it into four clusters at each iteration. In each cluster, the mean and variance of each RP defines a score which will be used to decide whether the cluster should be subdivided or not. Basically if the score does not satisfy a specific criterion (usually that it is not above a threshold) the cluster is divided again into subclusters and so on.

5. **Weighted clustering**: This mechanism defines a weighted connection as a similarity between two nodes. In this case it is assumed that close RPs receive similar readings from the same set of APs. Thus, the similarity is defined as a formula proportional to the inverse of the Hamming distance. Essentially the system chooses randomly a RP as cluster head (CH) and include all RPs satisfying that their similarity is over a predefined value (followers) in that cluster. Which actually means that a node might follow more than one CH. After all the clusters are defined a test is conducted in order to select the best node, that is to say the one having the least variance of FPs among all cluster members, as CH (might change the previous one). Finally the coarse localization is performed by selecting the cluster with the CH having the least distance from the on-line measurement and adding all the clusters sharing any RPs with the selected one.

6. **Spectral clustering**: Here the similarity measure is the pairwise cosine between the RP and the mean of the FPs in a cluster. There
are many possible applications. For instance, RPs are grouped into K clusters so that the similarity of RSS vectors is maximized.

7. **Layered clustering**: This is another procedure using AP coverage vector. The idea is a layered clustering of RPs based on their similarity with the on-line reading. Principally the system computes the Hamming distance between the on-line measurement coverage vector and that of each RP. Note that the coverage vector is the one defined previously in the first method (see 1). Next, the range distance is divided into K groups (set experimentally or by a training set) and the RPs are clustered with respect to their Hamming distance to the on-line measurement. Basically a RP is in a group if its Hamming distance corresponds to it. However, it might happen that an RP belongs to two successive groups in which case, the system will choose randomly one of them. Lastly, each group is weighted with the inverse of the average group Hamming distance. This way, all groups help during the fine localization step.

Interestingly enough this is not a coarse localization method but a group sparsity based localization one that combines coarse and fine localization in one step.

To sum up, these approaches exploit the off-line data and thus their complexity is not really a problem. However, most of these systems consider the statistical properties of FPs without taking into account the on-line phase. As a consequence, this might mislead the fine localization procedure. That is why binary AP coverage (see 1) and layered clustering (see 7) operate better in these scenarios.

### 1.4 Radio map construction

Several problems arise with large scale deployments. One of them is the collection of huge amount of received signal strength (RSS) data for high accuracy which leads to very high deployment costs. Additionally, the radio map changes over time and requires periodic calibration. That is why some researchers have focused on reducing data collection and on improving radio map generation. There are different new techniques to deal with these problems.
In this section these techniques will be presented and briefly discussed.

1. **Crowdsourcing-based data collection**: Many recent works have focused on crowdsourcing-based localization systems. These kind of systems ask users to voluntarily report their location and their Wi-Fi fingerprints (FPs). It means that the more users upload signals, the more the fingerprint database evolves. This is known as organic indoor localization (OIL) because it is said that the database organically evolves with input data. In essence, the users are helping updating the radio map. A slight challenge is the influence of feedback error or noise. Researchers have dealt with it by means of a clustering process that filters the wrong user input.
   
   This approach requires more development since the constant feedback to users might interfere in their experience. Another problem is the initial feedback data which carries too much noise because the users have not been accustomed.

2. **Implicit user participation**: The idea behind this technique is to transparently collect RSSI vectors. That is to say that the users help collecting the data through their daily life routines. For instance, there is a recent system called WILL [7] that records RSSI vectors and relative distances when the user is walking during the training phase using inertial motion sensors. Then, the system maps the corresponding RSSI vectors to the floor plan and builds the radio map.
   
   Another recent approach is called PiLoc [8] and it also uses motion sensors but provides more details in the indoor map construction.
   
   One limitation of works like WILL is that the collection of FPs relies on the step counter and the displacement measurement. Adding noise, the data might gather around a certain area affecting the quality of the survey. In real deployments the idea is to include some landmarks such as RFID tags or Wi-Fi for motion-sensor calibration, thus reducing the step counting error.

   Additionally, sometimes there are difficulties to build up the fingerprint database. But if the mapping relations between RSSIs and locations are obtained then just with a location fix from GPS it is possible to locate users based on RSS mapping scheme. An example of development that puts this idea into practice is called EZ [9] and it achieves to
reduce the survey cost while not requiring prior knowledge of the environment. Nevertheless, in a real scenario the location fix from GPS can be difficult to obtain. That is why there is an improved version of EZ called EZPerfect [10] that has been developed using labeled FPs. Ultimately researchers have build upon machine learning schemes. In essence these systems model the signal propagation for prediction at different indoor locations. As an example, the work called WiGEM [11] uses Gaussian mixture model and expectation maximization to estimate the target location and all signal propagation parameters. This way, RSSs at different locations can be predicted reducing the survey process.

3. **Use of partially labeled FPs**: Multiple studies have focused on the labeling process of FPs due to its difficulty in real scenarios. One interesting approach is HIWL which uses hidden Markov model (HMM) to classify unlabeled signal data. Initially there is a HMM training phase allowing to map the relationships in geographical and signal distribution. Then the system makes use of the map to match unlabeled FPs to their physical location. Although it may seem a simple system, HMM increases a lot the complexity and the data set which has to be large enough to ensure accuracy.

Another worth mentioning scheme is known as UMLI [12]. It uses clustering methods (see Section 1.3) to classify neighboring RPs having similar signal patterns. Basically using clustering methods enables to classify unlabeled signal data in locations with similar signals. One particularity of this system is the use of a hierarchical structure. That is to say, the procedure starts by classifying FPs to corresponding rooms and then comes the fine localization stage based on the coarse location result. The problem is that hierarchical localization increases the complexity of the system.

More recent developments have taken advantage of the correlation between labeled Wi-Fi FPs. For instance works like Co-Embedding [13] benefit from this idea reducing multi-floor survey cost. In essence, the system considers that in a building there are similar floor plans and that wireless signals are correlated. The algorithm behind first analyses the relation between FPs at different floors and later find locations of unlabeled RPs at other floors. It is clear that this practice is of no use for buildings with very different floor plans. Additionally it is quite
complex and the result does not provide with any error analysis.

As a conclusion, the techniques mentioned above can help with large data collection and map generation. Nevertheless the main issue still is the balance between cost and accuracy. Surprisingly enough, these systems achieve an accuracy lower than with traditional fingerprinting. Plus they often need a post-processing step for noise and errors which increases complexity and costs. That is why in scenarios with a dense number of visitors and high accuracy need, traditional fingerprinting survey is still used. However, in scenarios with a low number of users reduction algorithms may be employed to reduce deployment costs.

1.5 Density and weight estimation

One important issue regarding fingerprinting systems is the fact that fine localization accuracy highly depends on the distance between on-line measurements and received signal strength (RSS) radio map fingerprints (FPs). This means that an incorrect metric might cause biased estimation. To deal with this, researchers have developed different techniques to better exploit the off-line data and to ameliorate the comparison with on-line measurements.

The main idea here is to use weighting methods to increase accuracy of the estimation process. Nevertheless, other techniques do not assume that RSS follow a Gaussian distribution and focus on estimating the probability density functions in order to obtain the weights for location estimation.

In this section several density and weight estimation will be presented.

1. Kernel density estimation (KDE) method: Usual analytical assumptions on the prior probability on RSS fingerprints (Gaussianity) do not necessarily hold. Basically the problem is that parameter estimation do not capture empirical characteristics. The idea here is to estimate FPs distribution non-parametrically. KDE methods allow this by using a superposition of kernel functions centered around the FPs which can later be used for weight computation. For that matter, weights are obtained through the average normalization of the inner product between FPs and on-line measurements. In essence, this
metric measures the angles between the on-line measurements and the radio map FPs. It is important to notice that if the AP readings are correlated and the angle is small, this is not a representative metric.

2. **Principal analysis component (PCA) method**: This alternative approach consists in mapping the on-line measurements to the domain of its principal components (PCs). This transformation is achieved through concatenating the eigenvectors of the global covariance matrix of the FPs corresponding to the eigenvalues sorted decreasingly in a matrix called transformation matrix. Then, FPs and on-line measurements are transformed to the PC domain through the product with the transformation matrix.

3. **KL-divergence method**: The Kullback-Leibler divergence is basically a distance between two probability density functions, the on-line measurements one and the RSS fingerprints one, written as a kernel function. First, to obtain the probability density function of the online measurements the user has to remain at his/her location in order to get multiple measurements. After that, the KL divergence is computed and later combined with a kernel function allowing to yield weights for location estimation.

4. **Geometry-based localization**: There are some recent methods that exploit the geometry of the area to compute the weights for location estimation. For instance Tilejunction, Sectjunction and Contour-based trilateration. Those systems calculate the weights by solving a convex optimization problem with environmental constraints such as the presence of walls.

As a conclusion we have to say that the methods presented above in this section exploiting the density and weights are computationally complex. Additionally, estimating the density of FP distributions requires to adjust multiple parameters. That is the main reason why methods estimating weights are more appealing that the ones estimating FP densities.

1.6 **Additional recent approaches**

Previously the more important and interesting fingerprinting approaches for this specific report have been described. But there are many more tech-
niques that have not been presented. For instance, the sparse reformulation of the WLAN localization problem and the techniques for detection of outliers. Although these methods will not be used that much in the following chapters it is important to briefly mention them in order to complete this state of the art in fingerprinting-based indoor positioning systems. In this section several groups and types of techniques will be presented and briefly discussed.

1. **Exploitation of access points**: The complexity of indoor environments leads to several challenges associated with access points (APs). Basically there are three main problems: unavailability of APs, large set of APs and faulty APs. Here are some recent techniques to deal with it.

   - **Uniform AP selection methods**: These methods are told to be uniform because they select a uniform subset of APs for all RPs. Early studies have focused on selecting APs based on their signal strengths in the on-line phase. These are known as **Strongest APs** or **MaxMean methods** since the intuition is that strongest APs will provide coverage for most of the time and increase accuracy. Other researchers leaned towards the **Fisher criterion** which is a metric quantifying the discrimination ability of each AP across RPs and takes into account the stability of AP fingerprints. The criterion is based on the fact that APs with higher variance are less reliable than the ones distinguishing between RPs. A similar approach to the Fisher criterion is the **Joint Selection method** which instead of computing the differentiability of RPs with respect to the mean RSS, computes the mutual differentiability between RPs. Another idea is the **Group discrimination procedure** which selects a whole group of APs providing maximum discrimination instead of independent APs. **Information Gain (InfoGain) techniques** are based on the off-line criterion of selecting the APs with the highest discriminative power. Finally, there is the idea of **Entropy Maximization**, selecting the APs with the maximum entropy.

   - **RP-based AP selection methods**: Unlike the previous meth-
ods, these ones select a set of APs for each RP individually. One recent approach uses the Bhattacharyya Distance to measure the distance between the probability densities of the FPs of two APs. This way, it selects pairs of APs for each RP with the smaller distance.

The other procedure is similar to the last one since it measures the distance between the RSS fingerprints. It is called Information Potential and it uses a kernel function to calculate the distance and later assigns pairs of APs to the RPs.

2. Sparsity-based localization: On the one hand probabilistic approaches are computationally complex and on the other, deterministic approaches have a low accuracy. To face this situation, the sparse reformulation of the WLAN localization problem has been proposed. Here are some techniques solving this problem.

The main idea of this new model is that the localization problem can be interpreted as finding one location among all RPs corresponding to the closest to the user. For this matter, the location vector is transformed into a sparse vector which leads to an under-determined problem with generally infinite solutions. To solve it we might use Compressive-Sensing (CS) [14] which in specific conditions will give as a result a unique solution by using convex optimization and algorithms. This type of localization known as CS-Based localization employs algorithms such as greedy ones, iteratively re-weighted linear-squares (IRLS) and basis pursuit. LASSO-Based localization [15] is quite similar to the previous one but tries to avoid the strict conditions to get a unique solution and at the same time works better with noisy measurements. This is achieved through minimizing the norm of the location vector and its residuals.

Another recent approach are GLMNET-Based localization systems [15] which find a compromise between ridge regression and the LASSO method. Basically these systems consider the correlated predictors and find a sparse solution for the user’s location. Finally there is also Group Sparsity (GS)-Based localization [16] which does not follow the general procedure of fine localization in which the coarse stage finds a subset of RPs where the user is because it might lead to a wrong subset of RPs. That is why the systems employing this
technique use all clusters and weight them for the optimization phase.

3. **Assisted localization:** Many recent works have focused on techniques employing additional information from wireless device sensors to improve Wi-Fi fingerprinting systems.

- **Sensor fusion assistance for localization:** Wi-Fi networks are originally not designed for localization and in many cases they introduce distortions. That is why the use of sensors can help with this matter. For instance, *ambient lights and sounds* can help defining the area where the user is as if it was a coarse localization. Another idea is to benefit from *RSSI measurements* taken from other networks, like cellular ones, and use them in situations like poor Wi-Fi infrastructure deployments, weak signals and location ambiguities. Lastly, *RFID tags* can provide a completely independent location estimation from RSS fingerprints. Then, both location estimations are weighted and combined to get a final result. Note that here an additional infrastructure update is required.

- **Motion assisted localization:** There are some sensors called *Inertial Measurement Units (IMUs)* allowing dead-reckoning systems to estimate the change of the user’s position with respect to the past one. Some of such sensors are barometers, accelerometers, gyroscopes and magnetometers. In essence, they allow to collect user’s motion patterns from which it is possible to obtain valuable information like walking direction, detection and step counting. This information can later be used to help estimate the vicinity of the user’s location.

- **Landmark assisted localization:** Landmarks can sometimes help to detect certain locations in indoor environments that have identifiable signatures. There are two main types of landmarks: *seed landmarks (SLMs)* and *organic landmarks (OLMs)*. The first type can be associated to actual locations and the second type to sensory signatures. For instance, there is a recent work called UnLoc [17] which looks for certain structures in a building forcing the user to have predictable motion patterns. Another example
is known as SemanticSLAM [18] and it is interesting since it can recognize turns in corridors, classrooms and more.

- **Collaborative localization**: Another group of approaches focuses on sensors exploited for the nearby peers. The information can be used to extract the distance between wireless devices allowing to obtain relative locations or to improve the accuracy of the system. Multiple sensors can be used in order to find the distance between devices. For instance speakers and microphones can be used to transmit signals and calculate it which is known as **acoustic ranging**. Alternatively **Bluetooth** can be employed collecting FPs from devices with this technology in a database. Lastly magnetometers can help reduce the effects of magnetic perturbations, thus improving the accuracy of the system.

- **Opportunistic localization**: Unlike the above-mentioned approaches, this localization procedure does not require a dedicated infrastructure. The main idea is to maximize the exploitation of sensors and information when available. Basically, it extends the concept of landmarks to radio signals which is dynamic and complex.

4. **Outlier detection**: Sometimes APs do not work as expected due to various causes like obstacles, visibility or unavailability. Outliers occur when on-line measurements from an AP are different from any FP in the area.

- **Hampel filter**: The Hampel filter has been extensively used in statistical data as an on-line and off-line outlier detection procedure. It replaces the outlier-sensitive mean and standard deviation estimates with the outlier-resistant median and Median Absolute Deviation from the median.

- **Modified distance-based outlier detection**: This technique is a KNN modified version where the Euclidean distance between on-line measurements and FPs is done in an adjusted subset of APs. This way missing on-line measurements are excluded.

- **Sparsity-based outlier detection**: Sparse recovery methods discussed before (CS, LASSO, GLMNET and GS) can be em-
ployed in presence of outliers. They allow to estimate the corrupted APs jointly with the position.

5. **Heterogeneous devices**: Recently, one important issue that fingerprinting-based indoor positioning systems have to face is the heterogeneity of devices. The hardware differences between brands and models can degrade the accuracy of the system. In essence, wireless devices do not read equal RSS measurements partially due to the fact that they have installed different network interface cards (NICs).

Overall the main idea here is to compensate for hardware differences. Several techniques like linear regression, expectation maximization and neural networks are used to transform RSS data. Additionally the Pearson correlation can be employed to find similarities between RSS fingerprints and on-line measurements. Alternatively, normalization techniques and rank-ordering of APs may help with device-invariant FPs. Lastly, some works have used the signal strength difference (SSD) instead of the RSS to mitigate the effect of hardware readings from different devices.

6. **Energy efficiency**: As expected, indoor location based systems are similar to GPS in the sense that they are energy-consuming. The major issues when it comes to Wi-Fi fingerprinting are Wi-Fi scanning and data transmission.

- **Reducing scanning frequency**: The first idea that comes to mind to researchers is to control the scanning frequency. One way to deal with it is through *mobility suppression* which uses motion sensors to detect when the user is static. In such case the device suppresses the scan until the user moves again. Other options have focused in optimizing the Wi-Fi scanning according to the need of accuracy.

- **Reducing APs scanned**: Other techniques have leaned towards modifying the number of APs. For instance by only using the useful channels of APs. Alternatively, there is the possibility to use subsets of given AP lists to reduce on-line computation which is known as *dimension reduction.*
• **Replacing Wi-Fi with energy-efficient collectors:** Recently works have been developed on using new technologies to face with energy related problems. For example, there are two recent systems called *ZiLoc* \[19\] and *ZiFind* \[20\] using ZigBee (802.15.4) technology. One of the interesting things about this technology is the fact that it shares the same frequency channel, 2.4GHz, than Wi-Fi. Additionally, ZigBee has a network interface programmed to capture packets in adjacent bands allowing Wi-Fi fingerprints and on-line measurement through it. This can significantly reduce energy consumption.

As a conclusion it can be noticed that as seen in this section there are many works on fingerprinting techniques to face with all the challenges this technology brings. Besides, there are also some issues that have yet to be dealt with.
Chapter 2

Binary fingerprinting-based indoor positioning system

As discussed previously, fingerprinting (FP) allows the use of a variety of measurements for localization. The most common one is the received signal strength indicator (RSSI). In a FP system there are two different phases: the off-line and the on-line one (for details see Section 1.3). During the first step, measurements are collected and stored in a database which in general is laborious and time consuming.

This chapter focuses on reducing the quantization levels of the RSSI up to the binary state. Basically, the number of levels will depend on the conditions of propagation. Ultimately the goal is to develop a smart system able to adapt to these conditions as well as efficiently manage the available resources. This technique allows to place base stations (BSs) in an optimal way using concepts taken from error correcting codes theory.

As it will be seen throughout this chapter, the use of this technique reduces the use of memory. Additionally, the efficiency of the system from a computational point of view can be improved as well as the accuracy in some cases.
2.1 Binary fingerprinting technique

Throughout this report two main issues of fingerprinting techniques have been emphasized: the laborious and time consuming aspects of the off-line phase and the received signal strength indicator (RSSI) which can be inaccurate in indoor environments. To deal with these problems this chapter focuses on exploiting the quantization of the RSSI till to the binary level. Most of the information is directly taken from the two publications by M. Mizmizi and L. Reggiani already mentioned (see [2] and [1]). Throughout this section the technique will be presented and discussed.

Initially the system will have a set of $N_B$ static nodes or beacons equipped with Bluetooth technology (IEEE 802.15) that will help to localize the user in a limited area. These sensors normally receive the RSSI in a signal of 8 bits that goes from 30 dBm to -127 dBm. Using binary quantization there is a reduction of 7 bits that is to say 87.5%.

In a general fingerprinting-based system there is an important bottleneck when it comes to matching on-line and off-line FPs. Typically the system computes the Euclidean distance to check pattern matching of the $K_T$ vectors in the radio map. The binary fingerprinting system allows to adopt a different point of view. Here there are $\log_2(K_T)$ bits enumerating $K_T$ grid positions. Additionally their FP signatures are encoded in $N_B \log_2(L)$ bits with $L$ being the number of quantization levels. Since there is a binary quantization, there are only two levels, $L = 2$ and the beacons’ covered area is divided into two zones by means of a threshold $RSSI_{REF}$. To put it mathematically:

$$ r_i = \begin{cases} 
1 & \text{if } RSSI_{measured} \geq RSSI_{REF} \\
0 & \text{Otherwise} 
\end{cases} $$

(2.1)

To simplify the problem, let us consider grids of $2 \times 2$ cells (as in Fig. 2.1 (a)). The radio map of this first configuration, in Table 2.1, will allow to make some general observations and then proceed to formalize the construction of larger grids:
Table 2.1: Example of FP mapping (a)

<table>
<thead>
<tr>
<th>Fingerprint vectors</th>
<th>BS₁</th>
<th>BS₂</th>
<th>BS₃</th>
<th>BS₄</th>
<th>BS₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R₂</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R₃</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R₄</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.2: Example of FP mapping (b)

<table>
<thead>
<tr>
<th>Fingerprint vectors</th>
<th>BS₁</th>
<th>BS₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R₂</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R₃</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R₄</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2.1: Fingerprint grid with $K = 2 \times 2$ and corresponding covered areas. The triangles are the beacons: (c) is the covered area with the label channel model while (d) is an example of real coverage.
1. The use of the Hamming distance \( d_H \), which is the number of bits that are differing between two vectors, is of crucial importance for the following analysis. Essentially, the more distinct are the FPs vectors, the better will the performance and accuracy of the system be. Translated to the Hamming distance, the higher is the minimum Hamming distance \( d_{H,\text{min}} \) between the vectors, the better.

2. From the point above mentioned and observing the radio map, in Table 2.1, it is clear that some beacons like number 5 do not increase \( d_H \) between any of the vector and thus are useless.

3. Taking into account the observations made, the FPs vectors should be as in Table 2.2 which would require a configuration like in Fig. 2.2(b).

To deal with larger grids satisfying that \( d_H \) is at least equal to 1, the idea is to use an iterative procedure. In the following this procedure will be developed in two different scenarios. The first one considers that the RSSI depends only on the distance. The second one is more general and real due to the fact that shadowing and multipath are taken into account.

For this analysis let’s define some fixed parameters such as the square area \( A_0 \), beacons with A omni-directional antennas and B tunable transmission
power. Additionally, the iteration starts with $k = 0$. The following steps assign the binary signatures to the cells:

1. When $k = 0$, the first two beacons (i.e. two bits) are used to decompose the region $A_0$ into $N_{k+1} = N_1 = 4$ sub-regions $\{A_1(1), A_1(2), A_1(3), A_1(4)\}$ using the method seen in the configuration Fig. 2.1 (b). The result is shown in Fig. 2.2 (a) and $N_B(0) = 2$.

2. For $k > 0$, each of the $N_k$ sub-regions $\{A_k(1), A_k(2), ..., A_k(N_k)\}$ is divided into four uniform sub-regions just like in Fig. 2.2(a). Interestingly enough, there is a way to minimize the number of beacons consisting on reusing some beacons for adjacent cells sides. This way only four beacons can be placed in a cross configuration at the center of each $A_k(i)$ with $i = 1, ..., N_k$ (see Fig. 2.2) providing the additional two bits needed for the further division of each of the four sub-regions. It can be observed that this corresponds to a mirroring of the binary representation of the cells with respect to (w.r.t.) the sides occupied by the beacons. The total number of new sub-regions is $N_{k+1} = 4^{k+1}$ and the number of beacons at step $k$ is $N_B(k) = 2 + \sum_{i=1}^{n} 4^i$.

3. At each step two new bits are added. As stated before they correspond to the further binary division of the two dimensions of the area. The transmission range $R_k$ of the new beacons is reduced by a factor 2 and the cover area by a factor 4. So the new two bits at each step are associated to an $\text{RSSI}_{\text{REF}}$ that is evaluated according to the current $R_k$.

4. Check if the $\text{RSSI}_{\text{REF}}$ (or the required transmission range) associated to the beacons at the current step $k$ is compatible with the maximum transmit power of the devices. If it turns out to be compatible, increase $k$ ($k = k + 1$) and repeat the previous steps (from point 2) till to complete the binary labelling of the cells. On the other hand, if it is not compatible, then divide the area in four sub-regions and repeat the assignment (from point 1) independently in each of them.

The resulting design is characterised by the minimum number of beacons $N_{B,1}$ necessary for respecting the labelling without ambiguities in $K_T$ cells, i.e. with $d_{H,\text{min}} = 1$. With:
\[ N_{B,1} = 2 + \sum_{i=1}^{\log_4(K_T)} \]

In general the cells at minimum distance are also geometrically close since they are concentrated in the last step where each single bit flip change corresponds to passing half of the distance between two adjacent cells.

The focus now will be on the more general and practical scenario where the RSSI does not depend perfectly on distance because of the presence of shadowing or other propagation phenomena. When the RSSI has random components affecting its biunivocity with the distance, the grid construction and binary association described above do not respect the minimum Hamming distance equal to 1 anymore. This becomes more probable when the shadowing random component has low decorrelation distance w.r.t. grid size and large standard deviation (w.r.t. distance pathloss corresponding to grid spacing). Therefore, in presence of random RSSI components, the iterative procedure above mentioned is modified according to the following points:

1. Grid points and beacons positions are not modified and can be considered as the starting point of a finer placement and deployment optimization for a specific scenario. Although each particular shadowing realization could be subject to a specific optimal grid and beacon placement, the regularity and repeatability of this encoding offers implementation and design advantages.

2. For the \( k \)-th beacon, the number of grid points associated to the corresponding \( k \)-th bit is also maintained. This is strictly required since it is a fundamental step for maximizing the number of binary signatures associated to the grid points and minimizing the number of ambiguities; therefore, the first two beacons divide \( A_0 \) into \( N_1 = 4 \) sub-regions (not uniform in this case) and so on, iteratively.

3. For each beacon, \( RSSI_{REF} \) (or the transmit power) is chosen exactly in the middle point between the measured RSSIs of the two sub-regions; from a signal point of view, this choice minimizes the probability of error and, in case of no shadowing, it returns the same values corresponding to the previous sub-regions.

4. At the end of the encoding procedure, if some grid points have equal binary signature (\( d_{H,min} = 0 \)), new beacons are added for cancelling the residual ambiguities.
2.2 Channel model

Before continuing any further in the real case and simulations of the algorithm previously discussed, it is necessary to describe the channel model adopted for the analysis. This section will develop on the general channel model and on another one more specific to the project.

As explained before in the Introduction to Fingerprinting techniques (see Section 1.3), received signal strength indicators (RSSIs) are stored during the off-line phase. The most common model adopted for these RSSIs is the log-normal shadowing, that is to say:

$$RSSI(d) = A - L_0 - 10\alpha \log_{10}(d/d_0) + L_{SH}$$

where, $A$ is a constant given by the transmitted power and the antenna gains, $L_0$ is the average propagation loss at the reference distance $d_0$ (usually 1 m), $\alpha$ is the Path Loss Exponent (PLE), $d$ is the distance between transmitter and receiver and $L_{SH} \sim \mathcal{N}(\mu, \sigma^2_{SH})$ is the log-normal fluctuation due mainly to obstacles in the environment. It is important to bear in mind that $\alpha$ and $L_0$ depend not only on the relative position between a grid point and a beacon but also on the Line of Sight (LoS) and Non LoS (NLoS) conditions. In fact, each RSSI measured by the target during the on-line phase is affected by an additional measure variance depending on the environment changes, device orientation and device response. Such variance is probably the main issue in RSSI-based IPS. Clearly, the accuracy of the system decreases if the RSSI vectors are quite different than the ones collected during the training phase. Some measurement campaigns in indoor environments report a typical standard deviation within $2-2.83$ dB. Nevertheless, in this project because of the use of BLE sensors, the standard deviation will be approximated to $3.59$ dB. This RSSI variance in FP-based IPS is mainly due to two types of phenomena:

- Channel random fluctuations: That is to say slow and fast fading (shadow and multipath fading). There are several solutions to these problems but they are often not suitable for FP deployments based on BLE or systems reusing WLAN structures.
System impairments: For instance differences in target device types, user orientation, environmental changes between the two phases. The problem becomes even more serious in a pervasive environment with different users having different devices in different places (in a pocket, a bag or their hands).

As a result, the RSSI measurements are modelled as affected by an additional random log-normal component that is uncorrelated to the channel shadowing component. Then:

$$RSSI_{MEAS}(d) = RSSI(d) + W,$$

where $W \sim \mathcal{N}(0, \sigma^2_W)$.

### 2.3 Real scenario simulations and results

Before proceeding with the project, it is important to analyse some real scenario simulations of what has been done until now. This section will present some results taken from simulations in different environments with the help of a specific simulator.

The simulator is the one used in [1] and [2] for the simulations and exploitation of results. One of the main interesting features is that it allows to plot the cumulative density function (CDF) in function of the error (in meters) for both the binary fingerprinting method and an infinite bit case method. This is of great interest since it enables to compare the two methods and have an idea of the performance of the previously described system.

Another useful utility is the possibility to modify all the environment parameters as well as the dimensions of the cells and map. In fact, this is going to be used in the following to present the impact of some important parameters dealing with the environment in the performance of the system. Before presenting some graphical results, let’s go through the predefined simulation parameters. Regarding the environment itself, the cell side is two meters long and the whole area has sides of eight by eight meters. Which means that the basic environment has sixteen cells.

Now, let’s focus on the variables characterising the channel model mentioned in the previous section (see Section 2.2). The main propagation parameters
are initially set like in the following: -35 dB for the attenuation constant, 5 dB for the average loss at the reference distance, 2 as the path loss exponent (PLE), 1 meter as the reference distance, 1 as the number of measurements, 0 dB for the standard deviation of the received signal strength indicator (RSSI), 7 dB for the standard deviation of the RSSI during the on-line step and finally -95 dBm as the sensibility of the receiver.

It is interesting to start by evaluating the impact of the map dimensions in the system performance. It is shown in figures 2.3 and 2.4 with four and sixteen cells respectively. The result is clear: the smaller the map, the better the performance. It is logical taking into account what has been mentioned before about fingerprinting-techniques, they perform well in small areas.

Figure 2.3: Test simulation in a 4 cells environment.

Figure 2.4: Test simulation in a 16 cells environment.
Now, back to the original scheme, two simulations were run with sensibility values of -70 dBm and -150 dBm in order to evaluate the influence of the sensibility of the receiver. As it can be seen in the figure 2.5 and figure 2.6, the sensibility seems to affect quite slightly the performance of the method using an infinite number of bits. That is due to the fact that this method estimates the position by taking the minimum Euclidean distance between the on-line measurement and the RSSIs stored in the radio map which have now been modified due to the sensibility of the receiver.

Figure 2.5: Test simulation in a 16 cells environment and $P_{\text{min}} = -70 dBm$

Figure 2.6: Test simulation in a 16 cells environment and $P_{\text{min}} = -150 dBm$
The last parameter that has been tested is the standard deviation during the on-line step. Here again, two simulations were run with standard deviation (\(\sigma\)) values of 3 dB and 10 dB respectively. From the figures 2.7 and 2.8 it is clear that both positioning methods are strongly influenced by the value of this parameter. The higher it is, the bigger is the overall error.

Figure 2.7: Test simulation in a 16 cells environment and \(\sigma = 3dB\).

Figure 2.8: Test simulation in a 16 cells environment and \(\sigma = 10dB\).
Chapter 3

Optimization of the binary fingerprinting-based indoor positioning system

Until now, this report has presented reviews, some analysis and simulations of past and recent works. It is now time to investigate some variations and improvements of the design proposed in [1] and [2].

As mentioned before, the main novelty adopted in the proposed design is the use of Ternary labelling instead of the Binary one. That being said it also involves a new take on the position estimation algorithm.

In this chapter, the ternary solution will be introduced and explained in order to improve the overall performance of the binary fingerprinting-based indoor positioning system introduced in Chapter 2. Additionally, the position estimation algorithm will be modified and further developed. The combination of both novelties will be adapted to the already mentioned simulator allowing to test the new system.
3.1 Introduction to the optimized system

As anticipated in the introduction to this chapter, the main idea is to optimize the binary fingerprinting scheme by adopting a ternary representation for the on-line phase labelling process. This section will go through the details of such representation.

The representation of the received signal strength indicator (RSSI) quantized till to its binary level has already been developed during the last chapter of this report. However, let’s refresh the main idea which is to assign binary labels by means of a predefined threshold. To put it mathematically:

\[
    r_i = \begin{cases} 
    1 & \text{if } \text{RSSI}_{\text{measured}} \geq \text{RSSI}_{\text{REF}} \\
    0 & \text{Otherwise} 
    \end{cases} \tag{3.1}
\]

One problem regarding the above equation is that it is quite restrictive. In the sense that the result will be either be a ’0’ or a ’1’ without taking into consideration that some RSSI measurements during the on-line phase might be affected by the propagation environment, thus leading to erroneous decisions. Actually, the main conflicts will occur in the boundaries of the cells. For these cases, the old system used the above equation and assigned a bit that in reality possibly was not the correct one.

To deal with this matter, the main idea is to determine a region denominated as uncertainty region around the boundaries separating the cells. This way, whenever an on-line measurement is in this specific region, the system assigns it a ’X’ value to differentiate it from the other two cases. This new ternary label will later be exploited to increase the performance of the system by modifying the position estimation. It will be developed in the next section.

Although the values used for the ternary label might be chosen differently, this project establishes that the so-called uncertainty region will be represented by a ’0’, and the other values used for the label will be ’1’ and ’-1’ assigned according to the following new equation:
\[
    r_i = \begin{cases} 
    1 & \text{if } RSSI_{measured} \geq RSSI_{REF1} \\
    -1 & \text{if } RSSI_{measured} < RSSI_{REF2} \\
    0 & \text{Otherwise}
    \end{cases}
    \quad (3.2)
\]

It should be clear that in the above equation, \( RSSI_{REF1} \) will always be greater than \( RSSI_{REF2} \). The new label modifies the representation of the original map explained in Section 1 of Chapter 2 (see 2.1) which was like the one in 3.1(a) and now has the representation illustrated in 3.1(b). A graphical representation of the use of the thresholds for both methods is put together in 3.1(c) for the original method and 3.1(d) for the new one.

Figure 3.1: New representation (b),(d) and old representation (a),(c) of the system.
3.2 Modifications and positioning algorithm

The main objective of using a ternary labelling is to improve the overall performance of the system. In this section, the modifications with respect to the original simulator as well as the new positioning algorithm will be presented.

The original simulator did two different things for the position estimation of the binary case. First of all it computed the nearest fingerprints (FPs) by means of the Hamming distance. Then, in case there was only one nearest FP, the corresponding cell was directly assigned as the resulting position estimation. On the other hand, if there was more than one nearest FP, the system used the minimum Euclidean distance between the on-line received signal strength indicator (RSSI) measurements and the radio map to estimate the position. Since this last step does not exploit the labelling, it was decided to modify it and use the same method used for the ternary case explained in the following.

For the ternary case, the computation of the nearest FPs is slightly more complicated. For each ‘0’ in the ternary on-line RSSI label, the new algorithm has to evaluate both ‘1’ and ‘-1’ possible values. That means that it has to deal with $2^n$ possible nearest FPs. In order to evaluate if the fingerprint can be considered as valid, the algorithm performs the following steps:

1. Check if the FP belongs to the binary radio map.
   - In case it is true, then store that FP.
   - Otherwise, compute its nearest FPs using the minimum Hamming distance and store them. It should be noted that this implementation does not allow to store more than one copy of each FP.

2. Once all combinations have been tested, compute the position estimation with the stored FPs.

Now, the new method implemented to estimate the position is the following. The idea is to compute the centroid or geometric center of all the cells involved. These specific cells are found by linking the nearest fingerprints to their corresponding cells. Mathematically, this is the equation used by the algorithm:
\[
\frac{1}{N} \sum_{i=1}^{N} C_i = \frac{1}{N} \sum_{i=1}^{N} (x_i, y_i) \quad (3.3)
\]

Where \( C_i \) is the set of coordinates of the center of the cells. As previously stated, these cells are the ones matching with the nearest FPs computed using the new algorithm.

The expectation by using this technique is an improvement on the accuracy of the system, specially in the boundaries between cells. That is because now, when there is a measurement on the so-called uncertainty region, the system will compute the centroid between all the cells involved.

### 3.3 Numerical results and comparisons

The exploitation of the new method should increase the performance of the system. This section will use the modified simulator with the newer positioning algorithm and present some numerical results and comparisons.

After implementing all the modifications stated in the previous section, it is time to test several environments with different number of cells and beacons and check out the results. It is important to bear in mind that the simulator has the same predefined parameters as the ones described in Section 3 of Chapter 2 (see 2.3).

Before analysing the results, it is necessary to briefly explain the legend adopted in figures 3.2 and 3.3. The infinite number of bits case (in blue) is the old method that had infinite precision and used the minimum Euclidean distance to determine the position. All the red color curves are related to the ternary case. The partially optimized curve is the one corresponding to the use of the modified algorithm but without taking into account the use of the '0's in the label. The second curve speaks for itself, it is the one corresponding to the complete optimization. The last one, is one simple method that randomly selects one ternary FP from the nearest stored ones and assigns the corresponding cell as the position estimate. On the other hand, the curves plotted in pink are the ones corresponding to the binary case.
Figure 3.2: Test simulation with a 4 cells environment. Red curves correspond to the Ternary case and pink ones to the Binary case.

Figure 3.3: Test simulation with a 16 cells environment. Red curves correspond to the Ternary case and pink ones to the Binary case.
The comparisons between all the different methods in figures 3.2 and 3.3 show that the optimization clearly improves the overall performance of the system. Its cumulative density function (CDF) in both environments is similar to the ideal case and even better. It is also worth noticing that the ternary case using the new position algorithm, that is to say without taking into account the effect of the '0's, also improves the system. These two observations support the initial expectations: the optimization of the system increases its performance.

Despite the fact that the above figures clearly prove the effectiveness of the optimization, it is important to check the results in the uncertainty region where the new system is supposed to perform much better than the old one.

For this matter, the simulation of a simple real trajectory has also been implemented. The trajectory is represented by a simple linear function going from the bottom-left corner of the environment all the way to the top-right one. This way, it is expected to see an improvement nearby the center of the environment when approaching the influence of the thresholds.

As introduced before the trajectory tested has the appearance shown in figure 3.4, for the four cells environment and in figure 3.6, for the sixteen cells one. The results are plotted in figures 3.5 and 3.6 respectively. The curves represent the mean squared error (MSE) (in $m^2$) in function of the x coordinate (in m.). As expected, the adoption of the ternary labels improves the accuracy of the system and specially in the already mentioned uncertainty region. Furthermore, near the center of the cell, the position estimation does not improve significantly. Additionally it is interesting to observe that the MSE is maximum when the position is near the borders of the map. This is due to the fact that at the border the system will have the higher incertitude and thus the probability of error will approach its maximum.
Figure 3.4: Trajectory representation in a 4 cells environment.

Figure 3.5: The MSE as a function of the x coordinate.
Figure 3.6: Trajectory representation in a 16 cells environment.

Figure 3.7: The MSE as a function of the x coordinate.
Conclusions and future work

In the past years, the increase in demand of Location Based Services (LBS) has led the scientific community and companies to focus on the field of positioning systems. More specifically on Indoor Positioning Systems (IPS), which still lack of a global solution unlike outdoor positioning, which relies on the Global Positioning System (GPS) for almost five decades.

Many IPS have been developed using different technologies and several techniques. One commonly used system that stands out from the rest in terms of cost and performance is the so-called Fingerprinting technique combined with the Received Signal Strength Indicator (RSSI). There are many approaches and variants of this specific system (Chapter 1), trying to strengthen some of its weaknesses. From the review made in Chapter 1, it is clear that the off-line phase and the variance of the RSSI are some of the problems yet to be dealt with. In this context, the binary-based fingerprinting technique which quantizes the RSSI with a smaller number of bits (Chapter 2) seems to improve the performance of the system.

This Bachelor Thesis proposes an optimization to that particular technique (Chapter 3), in order to achieve even better results. These propositions are the adoption of a ternary-based fingerprinting technique as well as a new adapted position estimation algorithm. Basically, this new system exploits the uncertainty of the RSSI on-line measurements to establish a region where it is not clear which estimate position decision is the appropriate one to make. Chapter 3 presents all the above mentioned ideas, simulations and results. Moreover, the results turn out to be satisfactory and the accuracy of the system is increased with respect to the old one.

In this project only a specific optimization of the system has been pro-
posed. There is still an effort to be done: for instance, (i) optimizing the choice of the two thresholds used for the ternary labelling and (ii) analysing the complexity of the position estimation algorithm in order to find some possible simplifications in order to reduce the computational workload (for ternary labels, the algorithm has to compute $2^n$ possible combinations where $n$ is the number of uncertain measures). Furthermore it would be important to test the new system in a real life scenario in order to confirm the results presented here.
Ringraziamenti

Vorrei esprimere la mia gratitudine al mio relatore di Tesina Luca Reggiani del DEIB (Dipartimento di Elettronica, Informazione e Bioingegneria) al Politecnico di Milano per la sua guida durante questo percorso.

Ulteriormente, voglio ringraziare Marouan Mizmizi per il suo prezioso aiuto durante lo sviluppo del progetto.

I would like to express my gratitude to my Bachelor Thesis supervisor Luca Reggiani of the DEIB (Dipartimento di Elettronica, Informazione e Bioingegneria) at Politecnico di Milano for his guidance throughout this journey.

Additionally, I want to thank Marouan Mizmizi for his valuable help during the development of the project.
Bibliography


