Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Design and Implementation of an Adaptive Indoor Positioning System based on Received and Interpolated Signal Strengths

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Software Engineering & Internet Computing

eingereicht von

Patrick Favre-Bulle

Matrikelnummer 0426099

an der

Fakultät für Informatik der Technischen Universität Wien

Betreuer: Thomas Grechenig
Mitwirkung: DI Dr. Gerald Madlmayr

Wien, 12.04.2015

(Unterschrift Verfasser) (Unterschrift Betreuer)

TU WIEN

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Design and Implementation of an Adaptive Indoor Positioning System based on Received and Interpolated Signal Strengths

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Software Engineering & Internet Computing

by

Patrick Favre-Bulle

Registration Number 0426099

to the Faculty of Informatics

at the Vienna University of Technology

Advisor: Thomas Grechenig
Assistance: DI Dr. Gerald Madlmayr

Vienna, 12.04.2015

(Signature of Author)

(Signature of Advisor)

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Erklärung zur Verfassung der Arbeit

Patrick Favre-Bulle
Ottakringerstr. 149, 1160 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)	(Unterschrift Verfasser)

Abstract

Automated technology-assisted positioning and context-aware services have become available to the mass market in recent years due to the decreasing cost of able mobile devices. Although there are established solutions for large-scale positioning, like Global Positioning System (GPS), no generic solution for these services for indoor applications has emerged yet. Providers usually struggle with the diversity and complexity of radio wave propagation in enclosed settings.

This thesis proposes a novel approach to indoor positioning based on the development of a new system named Seasnips. It is based on received signal strength indicator (RSSI) with IEEE 802.11, and combines different techniques such as trilateration with received signal strength (RSS) and scene analysis based on radio maps from interpolated actual measurements. It also adapts its underlying radio propagation model from the statistical analysis of RSS readings from the used site. The aim of this system is to provide a powerful tool which automates most setup tasks, is easy to maintain and scales while maintaining accuracy on par with current scene analysis-based services. To familiarize the reader with these algorithms and methods, the first part provides detailed information about positioning techniques, investigates the usefulness of radio propagation models as well as RSS for positioning, and summarizes important related approaches. The design and implementation details of the system architecture and a fully functional prototype follow in the next chapters.

As part of a real-world evaluation, Seasnips was compared to a commercially available indoor positioning provider based on fingerprinting with Wireless Local Area Network (WLAN) in different settings. The results indicate that the positioning accuracy offers a performance comparable to a state-of-the-art solution while outperforming it in terms of maintainability and scalability.

Kurzfassung

Technologiegestützte Ortung und darauf aufbauende Dienste sind in den letzten Jahren dank massiv fallender Preise technischer Geräte leistbar für ein weites Publikum geworden. Es existieren bereits ausgereifte Lösungen im Bereich der globalen bzw. im Bereich der Outdoor-Ortung wie zum Beispiel das Global Positioning System (GPS). Jedoch hat sich noch kein vergleichbares Services für den Innenbereich durchgesetzt.

Diese Diplomarbeit stellt das neu entwickelte Innenraum-Positionierungssystem SEA-SNIPS vor, welches auf RSSI mit IEEE 802.11 basiert und Techniken wie Triliteration mit RSS und Szenen-Analyse von interpolierten Signalstärken von aktiven Messungen einsetzt. Das verwendete Radiowellen-Ausbreitungsmodell wird anhand der kontinuierlichen Aufzeichnung der Umgebung adaptiert. Das Ziel ist es eine Lösung zu finden, welche die meisten Aufgaben automatisiert und so die Benutzung und Wartung vereinfacht sowie gute Skalierbarkeit ermöglicht. Um einen Überblick über die verwendeten Methoden zu erlangen, stellt der erste Teil der Arbeit wichtige Positionierungstechniken und -algorithmen vor und erläutert den Nutzen von Ausbreitungsmodellen und RSS im Bereich der Positionierung. Das Design sowie die Implementierung der System-Architektur sowie des entwickelten voll funktionalen Prototypen wird in den folgenden Kapitel vorgestellt.

Im Rahmen der Evaluierung werden die Resultate von Feldtests des Systems mit einem kommerziell verfügbaren Positionierungssystem, welches auf Szenen-Analyse mit WLAN basiert, verglichen. Dabei wird gezeigt, dass die Genauigkeit der Positionsbestimmung in etwa mit den Leistungen des kommerziell erhältlichen Systems mithalten kann, jedoch bei Wertungen im Bereich der Wartbarkeit und Skalierbarkeit deutliche Vorteile aufweist.

Contents

List of Figures **ix**

List of Tables **xi**

1 Introduction **1**

1.1 Motivation . . . 3

1.2 Contribution and Research Questions . . . 5

1.3 Organization . . . 6

2 State of the Art **7**

2.1 Survey of Positioning Techniques . . . 7

2.1.1 Geometric Methods . . . 9

2.1.2 Scene Analysis/Fingerprinting . . . 13

2.1.3 Proximity . . . 21

2.1.4 Overview of Related Approaches . . . 22

2.2 Radio Propagation: Large Scale Models . . . 26

2.2.1 General Models . . . 28

2.2.2 Models for Outdoor Attenuations . . . 31

2.2.3 Models for Indoor Attenuations . . . 32

2.3 Analysis of RSSI Method . . . 36

2.3.1 Background . . . 37

2.3.2 Statistical properties of RSS in IEEE 802.11 . . . 42

2.3.3 Additional Properties . . . 44

2.4 Analysis . . . 46

3 Design and Implementation **47**

3.1 Design . . . 49

3.1.1 Architecture . . . 49

3.1.2 Components . . . 52
3.2 Implementation . . . 62
3.2.1 Embedded Devices . . . 62
3.2.2 Server . . . 65

4 Evaluation **75**
4.1 Observations with RSSI . . . 75
4.2 Methodology . . . 79
4.2.1 Comparative Product . . . 79
4.2.2 Method . . . 79
4.2.3 Setup . . . 81
4.3 Results and Discussion . . . 83
4.3.1 Experiment 1: Residential 65 . . . 83
4.3.2 Experiment 2: Studio 94 . . . 87
4.3.3 Summary of Results . . . 90

5 Conclusion and Outlook **93**
5.1 Summary . . . 93
5.2 Future Work . . . 96
5.2.1 General Enhancements . . . 97
5.2.2 Advanced Site Adaptation . . . 97
5.2.3 Algorithmic Enhancements . . . 98
5.2.4 Active Client . . . 98

Glossary **101**

Acronyms **107**

Bibliography **113**

A Propagation Reference Tables **125**

B API **129**
B.1 Active Node . . . 129
B.2 Server . . . 131

List of Figures

1.1 Illustration of indoor positioning 2
1.2 Illustration of RSS over a floor plan 3

2.1 Positioning techniques stack 8
2.2 Geometric trilateration 10
2.3 Geometric triangluation 13
2.4 Fingerprinting architecture 14
2.5 Fingerprinting reference point grid 15
2.6 Euclidean vs. Manhattan distance 17
2.7 Smallest polygon example 19
2.8 Proximity signal grid 21
2.9 Radio wave mechanics 27
2.10 Illustration of RaPSor ray tracing tool 31
2.11 Ericsson multiple breakpoint model 34
2.12 ITU model path loss plot 36
2.13 Antenna radiation signature 40
2.14 802.11b channels 41
2.15 RSS distribution 43
2.16 Linear regression plot of RSS standard deviation 44
2.17 Measurement of multiple APs concurrently 45
2.18 Impact of orientation with RSS 46

3.1 General SEASNIPS Architecture 50
3.2 Phases of SEASNIPS 50
3.3 Active node component diagram 54
3.4 Component diagram of the service control server (SCS) 55
3.5 Complete map generation procedure 56

3.6 Job schedule sequence diagram 56
3.7 How circles are drawn to a raster 59
3.8 Trilateration guessing process 59
3.9 Manual adjustment of the guessed active node (AN) Positions 60
3.10 Map generation process 61
3.11 Picture of the used access points (APs) 64
3.12 Overview of the server's UI 67
3.13 Example of RSS graphs as displayed by the UI 68
3.14 Example of RSS matrix as displayed by the UI 69
3.15 Example of distance offsets against target-distances 69
3.16 Example of an adapted propagation model 70
3.17 Example of distance matrix after an analysis 71
3.18 Example of the visual AN editor 72
3.19 Example of the user positioning user interface (UI) 73
3.20 Example of a Positioning with offset 74

4.1 Correlation of mean to sample size in RSSI. 76
4.2 Correlation of mean to standard deviation in RSSI. 76
4.3 Minimum and maximum values versus mean in RSSI. 77
4.4 Correlation of mean to median in RSSI. 77
4.5 Normal distribution indication in RSSI. 77
4.6 Example of frequency distribution from Experiment 1. 78
4.7 Example of frequency distribution from Experiment 1. 78
4.8 Example of a positioning with offset. 80
4.9 indoo.rs tools. 81
4.10 Floor plan of Experiment 1. 82
4.11 Setup of Experiment 1. 83
4.12 Floor plan of Experiment 2. 83
4.13 Setup of Experiment 2. 84
4.14 Trilateration estimation performance of Experiment 1. 85
4.15 Site evaluation of experiment 1. 86
4.16 Performance of site evaluation of Experiment 1. 86
4.17 Test locations of Experiment 1. 86
4.18 Accuracy details of Experiment 1. 87
4.19 Trilateration estimation performance of Experiment 2. 88
4.20 Site evaluation of Experiment 2. 89

4.21 Performance of the site evaluation of Experiment 2. 89
4.22 Test locations of Experiment 2. 89
4.23 Accuracy details of Experiment 2. 90

5.1 Summary of accuracy performance. 94
5.2 Interpolated setup efforts. 95

List of Tables

2.1 Performance comparison: SVM, KWNN, MLP, BAY 20
2.2 Overview of Related Approaches 25
2.3 Path loss exponent of the log-distance path loss model 29
2.4 Power loss coefficient of the ITU indoor model 35
2.5 Floor penetration loss factors of the ITU indoor model 35
2.6 dBm scale 38
2.7 RSSI percentage conversation 39
2.8 Example of WLAN survey data 42

3.1 Active node API overview 54
3.2 Example of a distance matrix 58
3.3 Used devices overview table 64
3.4 Server technology stack 66

4.1 Summary of the results. 91

A.1 Full List: Path Loss Exponent of the Log-distance path loss model 125
A.2 Full List: Power Loss Coefficient of the ITU Indoor Model 126
A.3 Full List: Floor penetration loss factors of the ITU Indoor Model 126
A.4 Average Floor Attenution Factor 127

CHAPTER 1

Introduction

Technology assisted positioning plays a major role in today's society. Since the Global Positioning System (GPS) [34] was made freely accessible with full accuracy about a decade ago and receiver-modules became small and cheap enough to be used in mobile devices, commercial and private application is now widespread. Although the current GPS generation and similar satellite-based positioning technologies achieve a precision-rating to a couple of meters, they fail in any non-open environment like tunnels or buildings. But exactly this requirement is essential for services like indoor navigation, which therefore needs other methods of positioning. Applications enabled by this technology include navigation-assistance in complex structures like airports or hospitals to context-aware services for commercial and educational use. Depending on the environment and the expected goals, the used means may differ considerably which will be directly reflected in effort and budget [109]. Most of today's mainstream and low-cost solutions use data from received signal strength indicator (RSSI) of signal-beacons to calculate the relative position to them [98] (see figure 1.1).

The major part of existing solutions as of 2015 use IEEE 802.11 Wireless Local Area Network (WLAN) in combination with RSSI and *scene analysis* (see fig 1.2), also known as *fingerprinting* (see section 2.1.2). The reason is simple: cost and availability with relatively low complexity by avoiding difficult problems like modeling radio wave attenuation (see section 2.2). In urban areas a vast and steadily increasing number of WLAN access points (APs) is deployed and common WLAN-enabled mobile phones (aka "smartphones" as they are called today) may act as clients. One of the origins of this technique can be found in the research of Bahl and Padmanabhan with their system from 1998, RADAR [4], which

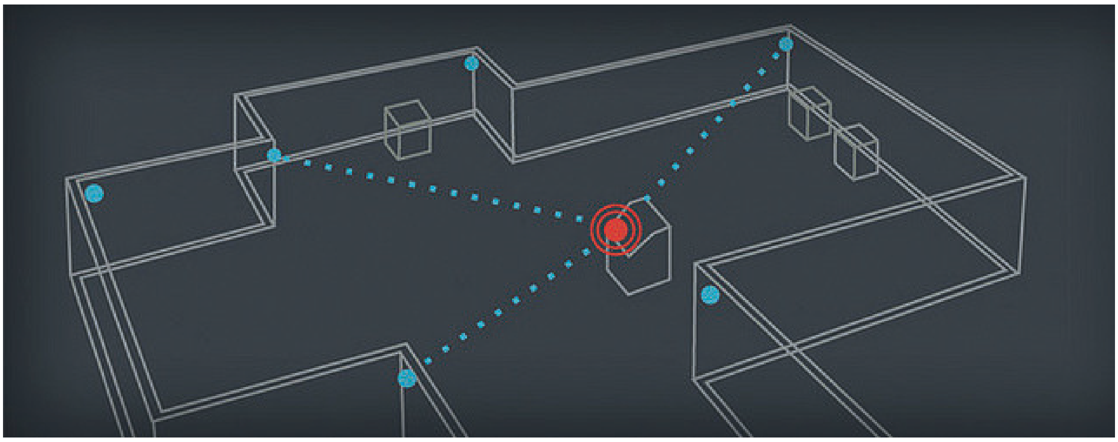


Figure 1.1: Illustration of indoor positioning [24]

used pre-IEEE 802.11 WaveLAN as wireless network (see section 2.1.4). Though lacking practicability at that time due to low availability of WLANs and capable mobile devices, it has now become one of the most used solutions. In its simplest form, *fingerprinting* does not even require a central administrative unit and mapping the area can be carried out with mentioned clients. The main drawback is its mediocre accuracy which can be expected to be around three to five meters on average. This of course limits its application to non high-precision positioning which is usually not a concern in simple user-centered services. Contemporary commercial indoor service provider incorporating *scene analysis* do exist. For example the Austrian-based company *indoo.rs* which, among others, offers implementations based on said technique. On the other hand, Google started providing content and application programming interfaces (APIs) for indoor maps with its popular product *Google Maps*, however, without a generically usable positioning system.

Another popular method is called *proximity* (see section 2.1.3) which either uses a known grid of receivers or emitters and estimates the relative position of the user to this grid with different methods and algorithms or just indicates if a user is near a certain beacon. It usually needs additional hardware to work and the achieved accuracy is highly dependent upon used setup. A current commercial product that uses this approach is Apple's propriety Bluetooth (BT) 4.0 system *iBeacon*.

Approaches that use other techniques, like geometric positioning, are quite uncommon for indoor positioning. They do play a major role in large area navigation like the above-mentioned GPS which incorporates trilateration with time of arrival (TOA) (see section 2.1.1) but typically are not applicable for the smaller scale. For example TOA for GPS requires high precision, atomic-clock synchronized and time-dilatation adjusted

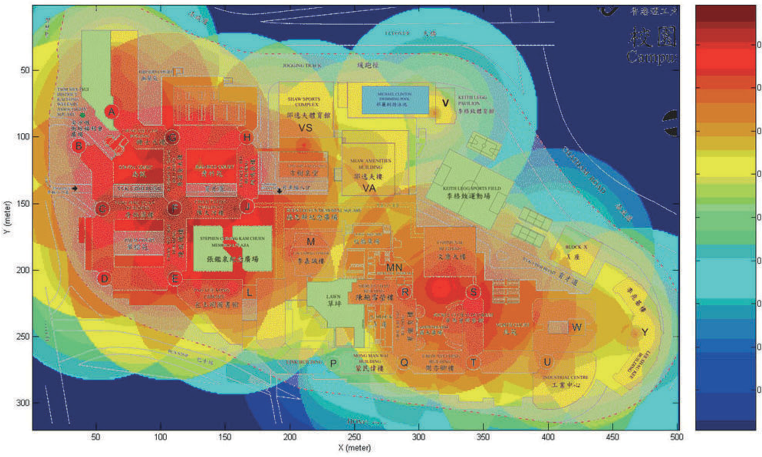


Figure 1.2: Illustration of received signal strength (RSS) over a floor plan [111]

time measurement which has to be precise to the nanosecond when accuracy should be in a reasonable range (i.e. under several hundred meters). Only specialized expensive hardware is able to fulfill these demanding requirements. Though attempts have been made to incorporate this approach with WLAN [54] it still had to be upgraded technically since the network interface controller (NIC) level responses are not under a millisecond and therefore about six orders of magnitude too slow .

Most of these systems are rated upon highest achievable accuracy and, secondary, cost. While these are important metrics, properties like robustness, scalability and maintenance often play a more important role in indoor services for large commercial applications such as navigation in train stations or large congress buildings than the aforementioned because every system has to be manually tailored to the environment. As of now, no generic system exists that is comparable to GPS and works without specific configuration for the majority of buildings and places. Research and available products often fail to address these issues with the needed intricacy, one major reason for the currently low penetration of location-based indoor services.

1.1 Motivation

As one of today's most important techniques for indoor positioning, scene analysis is especially prone to degradation of quality over time, requiring high maintenance to keep the service usable. Since it mainly uses the provided radio wave environment originating from WLAN APs either deployed by the provider himself for covering the area with wireless Internet and local area network (LAN) access or random emitters positioned by third parties in the near vicinity, the stability of these APs can not be guaranteed. Even devices owned by a provider may be intentionally or unintentionally moved if not fixated. This of course skews the measurements lowering accuracy consistently. On top of that, without built-in quality control it is not possible to check easily whether the service's precision is under a certain threshold. Therefore, consequent testing would therefore be needed to see if a re-measurement is required on top of the effort of the maintenance itself. For example the largest hospital in Austria, the Wiener Allgemeine Krankenhaus (AKH), provides a usable floor space of roughly 345 00 m^2 [110]. If only half of this area is mapped and a reference point is only every 16 m^2 and a single reference measurement takes about 20 seconds it would take around seven man-days to accomplish this. This shows the importance of scalability and robustness in such scenarios, because here the cost of implementation would be low, but the cost of maintenance unreasonably high.

Personnel responsible for measuring additionally has to be specifically instructed of the procedure to be able to do it properly. For example the effect of the human body on radio waves [35] has to be acknowledged to take special care in keeping a consistent orientation throughout the phase. Also the exact positions as marked on the map have to be assumed. Errors start to build up from this early stage. Even if these preparations are executed flawlessly, the fact that it only represents a snapshot of current properties in this site cannot be dismissed. Bahl et al. [5] show that radio wave propagation differs significantly depending on many factors, like present people or open windows and doors (see also section 2.2).

The problem with the other currently popular method, *proximity*, is that it cannot be generically applied to large areas. Each setting has to be pondered between the intended use, accuracy and available resources. It is therefore better suited to indicate "hot-spots" than to provide a general positioning system, although the latter is also possible with a grid layout. Here the degradation of quality is negligible because of facilitating an autonomous wireless system (e.g. with BT or radio-frequency identification) that is not dependent on external changes. Unfortunately manual maintenance is still required for recharging mobile devices or replacing defective ones. The difference is the added complexity in the design phase. Experience and basic knowledge of the behavior of the used beacons is needed to plan a system that meets the requirements. It also puts a bigger strain on the initial hardware budget, which makes it unattractive to use for nonrecurring brief events.

Generally, prior research mostly focused on the accuracy and precision of respective approaches and ignored administrative factors. Hardly any work investigates how practical their method would be to deploy and maintain in different scenarios. It appears that exactly these research questions in the field of indoor positioning should be investigate further. It has been shown numerous times that systems of nearly arbitrary precision requirements can be achieved using different methods which are often involved and complicated or sometimes not feasible in practice. Sophisticated procedures should be automated if possible and complexities of certain mechanics should be solved for the user without the need to be an expert in the field.

On the technical side, systems that use radio attenuation models as distance estimation usually choose between one of the well known models provided in literature (see section 2.2.3 “Models for Indoor Attenuations”). Although they usually provide some customization with parameters derived from empirical observations, they often are very general, resulting in only average accuracy. Researchers often try to compensate by adapting these models to their specific site properties though this is obviously not a feasible approach for practical deployment. It is a topic of ongoing research to find the best approximation of radio wave attenuation in confined spaces to derive distance, for example. Solutions vary between models defined by a simple path loss formula to sophisticated ray tracing algorithms (see section 2.2.1). A generally accepted and practical solution for RSS has yet to be shown.

1.2 Contribution and Research Questions

This thesis presents a novel concept for an indoor positioning system which was named *Site Evaluating Active Sensor Network Indoor Positioning System* (SEASNIPS) . It uses stationary active sensors unlike most IEEE 802.11-based systems, to aid in its process. The aim is to provide a design that is scalable, robust and requires low maintenance with keeping most process automatic while continuously adapting to the changes of the specific site. To achieve this, multiple techniques are combined which are among others: distance measurement with RSSI, node position estimations with trilateration and user tracking with simulated *fingerprinting* based on data from a site-adapted propagation model. The main components is a central server which administrates the system consisting of so called active nodes (ANs). These are just regular consumer-grade WLAN access points (APs) provided with additional custom software. They are used as WLAN sensors for measuring the surrounding signal environment. They can be either upgraded from already deployed APs or added as needed. The usage of active sensors for a continual site survey differentiates this system to most presented in literature. A fully functional prototype was implemented for evaluation purposes. Using this prototype, the following research questions are investigated in this thesis:

- Is RSSI with IEEE 802.11 a reliable and stable measurement?
- Does trilateration with RSS as distance estimation present a useful procedure for setting up the initial positions of all nodes?
- How does the accuracy and precision of a system using interpolated signal strengths inferred from actual measurements fare against current commercial products based on scene analysis in different situations?
- Does the presented approach result in a system with lower maintenance and higher scalability compared the systems currently used?
- For which applications and scenarios could such system be used?

The metrics used in research to review indoor systems is discussed in detail in chapter 4 “Evaluation”.

1.3 Organization

The necessary background information and state-of-the-art of current position techniques are discussed in a survey in chapter 2 “State of the Art”. This chapter also shares some insight of the basic concepts of radio propagation models used in the design and gives a detailed summary of important points when working with RSSI. Also related works proposing different indoor positioning systems are evaluated. In chapter 3 “Design and Implementation” the used scientific methods and design of SEASNIPS is illustrated following a section about details of implementation. The results of the performance can then be examined in chapter 4 “Evaluation” where also insights in the observed behavior of RSS in WLAN are discussed. Lastly, chapter 5 “Conclusion and Outlook” shares some ideas in which fields the system may be enhanced to further improve its performance.

CHAPTER 2

State of the Art

This chapter provides background and the *state of the art* of topics relevant to indoor positioning as well as related approaches in published literature. The first section provides a thorough survey of important positioning techniques that may be applied for indoor navigation which concludes with a discussion of three distinct approaches for indoor positioning systems. The next section describes the mechanics of radio propagation models and presents some important ones which are also later used in the implementation. The section about received signal strength indicator (RSSI) explains the basics about Wireless Local Area Network (WLAN), the used unit and statistical properties of said measurement.

For easier consumption, the related work from publications are interwoven with the basic background information. This is to not split the information about the same topics into two sections. This method especially was applied to section 2.1 “Survey of Positioning Techniques” and section 2.3 “Analysis of RSSI Method”.

2.1 Survey of Positioning Techniques

In the following sections different methods and processes will be presented which can be used to estimate the position of a mobile unit in an environment. Each of those techniques will be examined for their basic properties, procedure and specific appliances and value for shorter range indoor positioning. For brevity, techniques that lack any relevance to that topic are omitted so the presented list is not exhaustive. The section concludes with an overview of related approaches, applying some of the presented methods and algorithms in different ways.

In general, the basic techniques are either geometric, proximity or scene analysis. Geometric methods are mainly trilateration, which operates on distances, and triangulation which uses angles to estimate a location in different variations [60, 102]. These are well researched methods of mathematics that have applications in many technologies nowadays (e.g. Global Positioning System (GPS) [42]). Proximity is a simple concept that can easily be applied to indoor navigation applications [14, 64] but due to the fact that it requires additional hardware and setup, scene analysis based approaches are currently mainly adopted for low-cost indoor location-services [37, 59, 91, 113]. While not new, only the ubiquitous presence of WLANs in practical every building and the embedding of this technology in most mobile phones makes it a viable and easy solution deployable in most instances usable without any specialized hardware. This method, which is also referred as *fingerprinting*, also maintains a reasonable accuracy by mitigating impacts of imprecise radio propagation models although it lacks in scalability and degrades over time without maintenance.

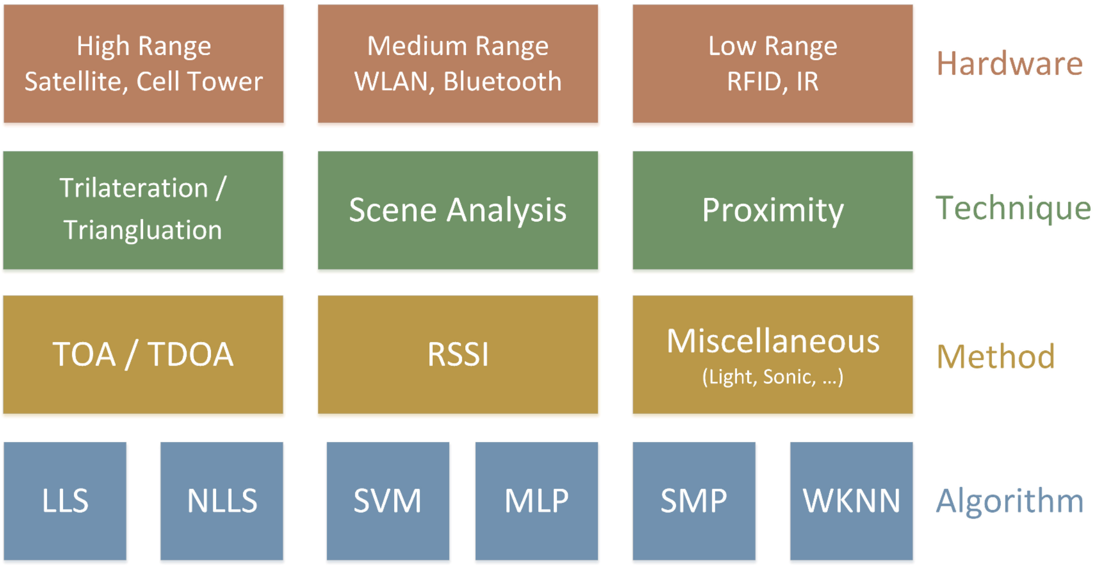


Figure 2.1: An overview of the positioning stack, which techniques methods and algorithms are relevant for positioning [25]

2.1.1 Geometric Methods

Methods that use geometric properties of triangles or circles to calculate the target location. They either utilize measured distances or angles for their respective process. Some of these techniques are known for centuries. For example triangulation is followed from the work of Willebroard Snell in the 17th century which was used for large-scale measurements and cartography [41]. Through evolution of technology, the methods discussed here see new applications not possible half a century ago. For simplicity only two-dimensional geometry is considered in-detail, although the methods works analogous in a three-dimensional space.

Trilateration

Lateration techniques estimate the position of an object by measuring its distances from multiple reference points. Instead of using the physical distance directly, received signal strength (RSS), time of arrival (TOA) or time difference of arrival (TDOA) is usually used to derive the data [97].

Algorithms There are different methods and algorithms to solve trilaterations. This is a none exhaustive list of a few important ones:

Trilateration in 2D Based on the equation for a circle

$$(x-h)^2 + (y-k)^2 = r^2$$

where h and k are x and y coordinates of the center of circle, x and y are where the circle has intersections with x-axis and y-axis and r is the radius and the following problem:

> Point a,b and c with distances to a target t which are r_1, r_2 and r_3 where the position of t i.e. the x- and y-coordinates t_x and t_y have to be found

and the assumption that a is at the origin $(0,0)$, b is at $(k,0)$ and c at (i,j) (see figure 2.2), we get the equations for the three circles [78]:

$$\begin{aligned}
r_1^2 &= x^2 + y^2 \\
r_2^2 &= (x-k)^2 + y^2 \\
r_3^2 &= (x-i)^2 + (y-j)^2
\end{aligned}$$

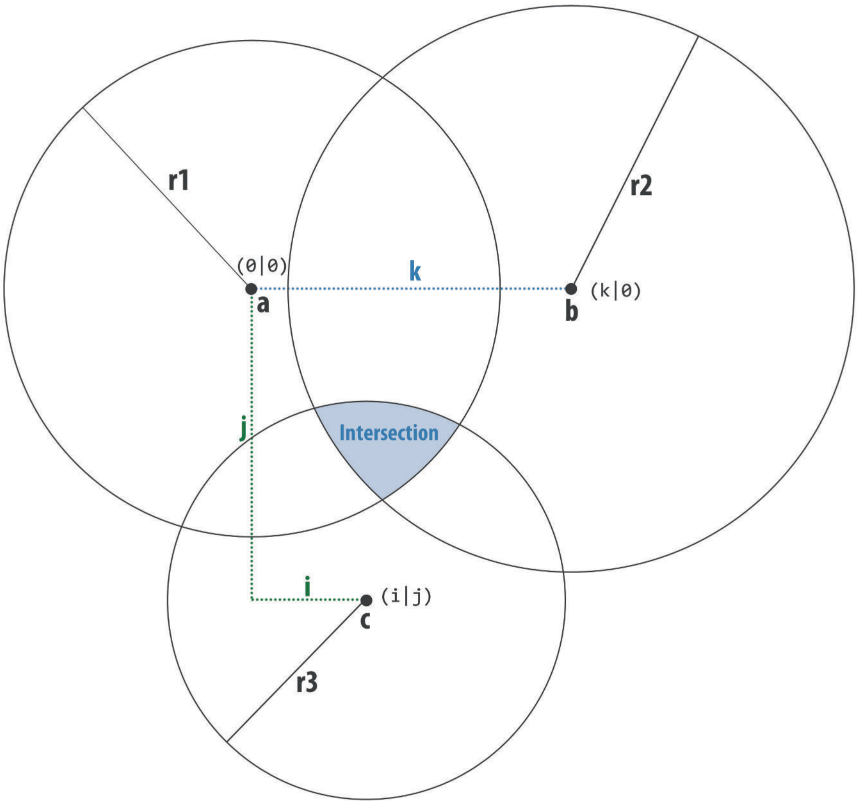


Figure 2.2: Example of a simple trilateration for Points a,b and c with estimated distances r_1,r_2 and r_3 to a receiver with unknown position which is calculated to be in the intersection with x-positive right and y-positive down.

We need to find a point at (x, y) that satisfies these equations which can be solved for x and y:

$$x = \frac{r_1^2 - r_2^2 + k^2}{2k}$$
$$y = \pm\sqrt{-\frac{r_1^2 - r_2^2 + k^2}{2k} + r_3^2 + 2yj - j^2}$$

y can be expressed as a positive or negative square root therefore the problem can have zero, one or two solutions. For negative values, the point is outside of the circle and no solution exist; if the circles touch at one point y is equal to zero, otherwise y is greater than 0.

Linear Least Square Method This algorithm is not very accurate also merely provides an initial position which can be used afterwards by other positioning algorithms. Therefore it can be disregarded as a stand-alone algorithm for indoor positioning [54,96].

Nonlinear Least Squares (Newton) The nonlinear least squares algorithm employs an initial estimation of the position which can be provided by a linear least square calculation and minimizes the sum of the squares of the errors in the distances. This is a fairly common problem in applied mathematics for which various algorithms can be applied [54,60].

Distance Calculation Apart from the different algorithms for trilateration there are multiple approaches for calculating the distances which can apply in different scenarios. The following describes some of the most important ones:

RSSI Based RSSI uses the attenuation of emitted signal strengths in combination with a radio propagation model to guess the length of the path the radio wave has traveled. RSS is basically the received signal path-loss due to propagation. It can be applied where line-of-sight path to the emitter is not available. Theoretical and empirical path-loss models are used in general and empirical models are mainly utilized for indoor environments specifically. Because of the complicated properties of radio waves (e.g. diffraction, absorption and shadowing) path-loss models do not always hold and can degrade the consistency of the data severely (see also section 2.2). Due to the fact that this is especially prevalent in indoor environments this method is seen mostly unfit in current research in combination with trilateration [69,70]. For an in-depth discussion about RSS see section 2.3.2.

Time Based TOA is the travel time a radio signal needs to propagate from a transmitter to a receiver. For TOA based systems the one-way propagation time is used to calculate the distance. Radio waves travel with the speed of light c in vacuum [12], but this deviates in accordance with the material which is utilized as physical media (for example c is about 90 km/s slower in air) which has to be taken into consideration.

$$distance \approx c \times time \tag{2.1}$$

For this method to work reliably the clocks of all participants have to be precisely synchronized then a time-stamp can be sent which the receiver can use for its distance calculation. This represents the biggest challenge with this method since an offset of simply 10 ns means a deviation of $\approx$ 3 m thus the synchronization process usually has to be built into the design of the system. For example GPS utilizes satellites that are synchronized with atomic clocks and corrected for time dilatation because of different gravity [2], but receivers usually lack any sophisticated time keeping function and are therefore less stable. To mitigate this fact satellites additionally transmit their current time which will be used to calculate the position. This also means that the minimum of four satellites must be in view instead of only three needed for trilateration because of time synchronization [42].

In comparison TDOA uses the difference in times at which a single signal arrives at three or more nodes or the difference of multiple signals from a single transmitter arriving at a another node lacks any complicated synchronization requirements [12]. The former is used e.g. in cellular networks because a mobile phone is synchronously connected to multiple base stations [97]. This is known under the term uplink time difference of arrival (U-TDOA) where it achieves real world accuracies of 100 m or better in 67 % of cases according to the research [116] of Yilin et al. Since the positioning process is done on the network side and the mobile phone does not take any part in the calculation process, it is supported by most devices and therefore also plays a role in emergency situations. Assisted Global Positioning System (A-GPS) is the latest iteration of this technology which combines GPS navigation with the aforementioned method but mobile phones have to support this technology in hardware [39].

Due to the fact that these properties TOA/TDOA is less than ideal for indoor positioning:

- the physical media radio signal travel in buildings vary greatly so the correct velocity can only be guessed making it more suitable for instances where every node is in a line-of-sight path
- precise time synchronization is needed by the system which cannot usually be provided with additional effort if not designed for it

Despite these drawbacks there is research and applications for utilizing TOA with WLAN. In the work of Izquierdo et al. [54] a procedure is proposed that uses Round Trip Time (RTT) for its TOA measurement but it is dependent upon an external hardware subsystem that is connected to the WLAN since software calculation would be in the magnitude of several milliseconds which is to slow. This is further improved by Ciurana et al. [20] with the use of a *Kalman Filter* [54].

Triangulation

Angle of Arrival (AoA) uses the capability to sense the direction from which the signal is received. This requires dedicated hardware like an antenna array, or several ultrasound receivers [83]. The location of the desired target can be found by the intersection of several pairs of angle direction lines, each formed by the circular radius from a base station to the target. The location of at least two points and angles to the target are required to find a location in two-dimensional and whereas in a three-dimensional system three points and their angles have to be known. The base formula for the calculation is (see figure 2.3)

$$d = m\frac{\sin\alpha\sin\beta}{\sin(\alpha+\beta)} \tag{2.2}$$

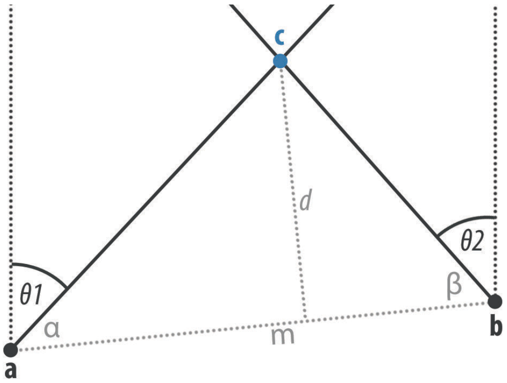


Figure 2.3: Example of a simple triangulation with 2 known points a and b with known angles to point c from the same base angle.

Triangulation requires relatively large and complex hardware and does not work well with moving targets. For accurate positioning, the angle measurements need to be exact, but this may be limited by shadowing, by multi path reflections arriving from misleading directions, or by the directivity of the measuring aperture [73] making it mostly unsuitable for most indoor positioning.

2.1.2 Scene Analysis/Fingerprinting

Scene analysis, or fingerprinting, refers to an algorithm that first collects data from defined points so they can be later compared to live measurements. This is also called the *offline* training phase and the *online* determination phase (see figure 2.4). The goal is to build an extensive database with fingerprints which can consist of measurements from various types of sensors ranging from altimeter through air pressure to thermometers, but in most solutions RSSI of wireless networks is the main indicator and what is focused on here [59,89]. It is also the basis of many currently used indoor positioning systems [16,58].

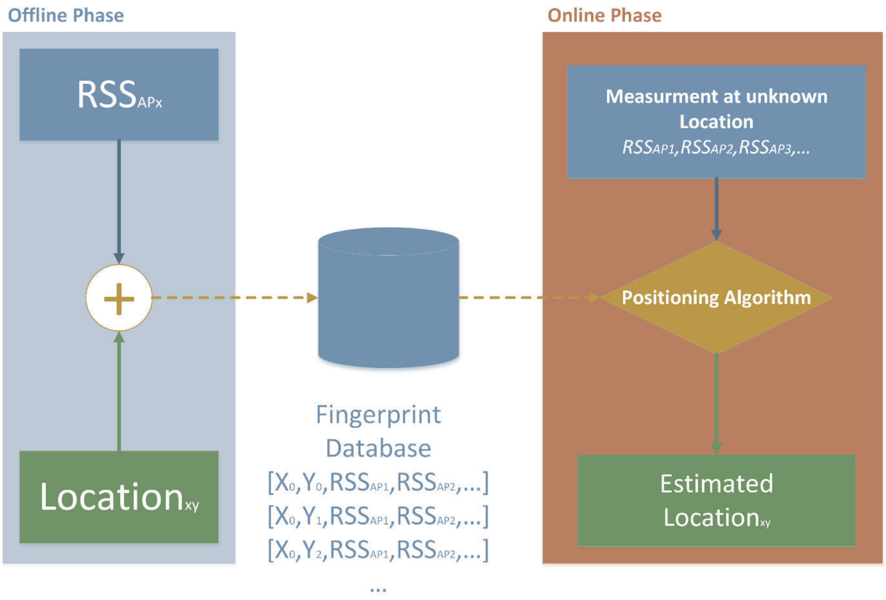

Offline Phase
RSS_{APx}
+
$Location_{xy}$
Fingerprint Database
$[X_0,Y_0,RSS_{AP1},RSS_{AP2},\ldots]$
$[X_0,Y_1,RSS_{AP1},RSS_{AP2},\ldots]$
$[X_0,Y_2,RSS_{AP1},RSS_{AP2},\ldots]$
...
Online Phase
Measurment at unknown Location
$RSS_{AP1}, RSS_{AP2}, RSS_{AP3},\ldots$
Positioning Algorithm
Estimated $Location_{xy}$

Figure 2.4: General outline of the fingerprinting algorithm [113]

The offline training phase starts with the definition of reference points in the given setting which are called location candidates in the online phase. Usually these are laid out as a grid with slight deviation conforming to the physical condition (walls, furniture, etc. see figure 2.5) or satisfying specific requirements for the positioning system. The resolution of this grid defines the maximal achievable average accuracy but also alters the effort needed to complete the training phase.

In wireless networks, access points (APs) or base stations are usually fixed positioned devices that emit a radio signal which in turn will be used as RSS for fingerprinting. So for every reference point the tuples of all RSS measurements from all APs in range are saved into an offline training database which represent the "*fingerprint*" of this location. This requires a certain minimum coverage of radio signals in the mapping area for accurate predictions. Special care has to be taken into consideration that measurements are consistent and representative; see section 2.3 for an in-detail discussion.

During the online phase a client with unknown location surveys the RSS of all base stations in range and uses it in combination with a positioning algorithm and the previously

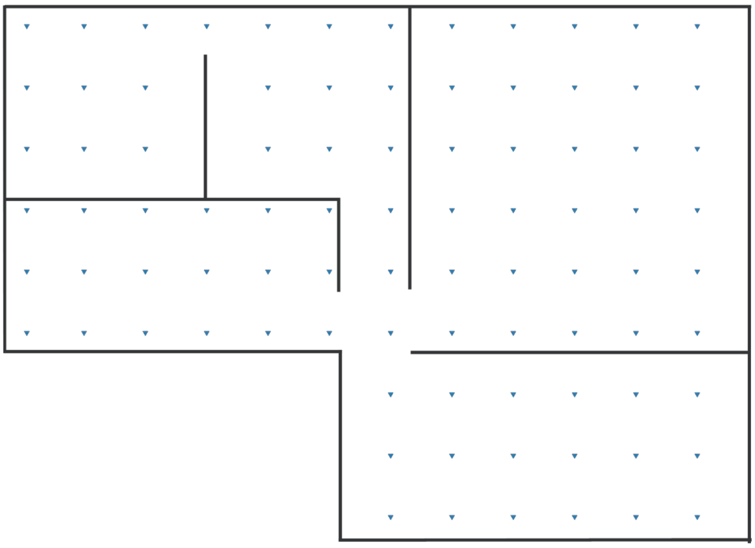


Figure 2.5: Fingerprinting reference points in a grid pattern

collected measurements to estimate its most likely location among all candidates. There are several possible types of algorithms that can be used for positioning:

Probabilistic Method

The basic idea is to find the location L with the posteriori probability given an RSS vector $\vec{s}$ measured in the online stage. It is assumed that RSS readings are distributed similar to *Gaussian* probability distributions [97]. Suppose there are n locations $L_1, L_2, L_3, \ldots, L_n$ the following *decision rule* can be obtained which selects the L_i where its probability is the highest [73]

$$\text{Decide } L_i \text{ if } P(\, L_i \,|\, \vec{s}\,) > P(\, L_j \,|\, \vec{s}\,) \quad \text{for } i, j = 1, 2, 3, \ldots, n \quad i \neq j$$

From *Bayes' theorem*

$$P(\, L_i \,|\, s\,) = \frac{P(\,\vec{s}\,|\, L_i\,)P(\, L_i\,)}{P(\,\vec{s}\,)} \tag{2.3}$$

the posteriori probability $P(\, L_i \,|\, \vec{s}\,)$ is the combination of likelihood $P(\,\vec{s}\,|\, L_i\,)$, prior probability $P(\, L_i\,)$, and observed evidence $P(\,\vec{s}\,)$. Since $P(\,\vec{s}\,)$ is the same for one positioning attempt, and the prior probability $P(\, L_i\,)$ that a mobile user locates at a specific point L_i is assumed to be the same for all reference points. The comparison of the posteriori probability could be considered as the comparison of likelihood:

$$\text{Decide } L_i \text{ if } P(\,\vec{s}\,|\,L_i\,) > P(\,\vec{s}\,|\,L_j\,) \quad \text{for } i,j = 1,2,3,\ldots,n \quad i \neq j$$

The likelihood of each location is assumed to be the *Gaussian* distribution. Thus the means μ and standard deviations σ of each location candidate could be calculated from the sample data. The signal emitting base stations in the target environment are assumed to be independent from each other, so the overall likelihood of one location candidate can be calculated by directly multiplying the likelihood of all m base stations

$$P(\,\vec{s}\,|\,L_i\,) = P(\,s_1\,|\,L_i\,) \times P(\,s_2\,|\,L_i\,) \times P(\,s_3\,|\,L_i\,) \times \cdots \times P(\,s_m\,|\,L_i\,) \tag{2.4}$$

At last we calculate the estimated location (x, y) as the average of the coordinates of all locations

$$(x,y) = \sum_{i=1}^{n} \Big(P(\,L_i\,|\,\vec{s}\,)(x_{L_i}, y_{L_i}) \Big) \tag{2.5}$$

by utilizing their posteriori probabilities as weights [72, 106]. This also has the advantage that not only the discrete location candidates can be picked but also depending on the calculated weights any location in between.

Kontkanen et al. proposes a Bayesian-network-based system in their work [65] to address other probabilistic modeling techniques for location-sensitive applications in wireless networks that involves active learning, error estimation and tracking-assisted positioning.

Deterministic Methods

In contrast to probabilistic models where states are not described by unique values, but rather probability distributions, deterministic techniques use uniquely determined parameters of the states of measured RSS vectors from APs. In the following different methods will be discussed that fall in this category.

Nearest Neighbor The nearest neighbor method calculates the distances between the live RSS vector of the online phase and the database entries of each reference point (see eq. 2.7). The location with the minimum distance is the Nearest Neighbor and the likely (x,y) location [113].

Usually Euclidean distance or Manhattan distance is used in such approaches. The Euclidean distance is the direct shortest path between two points in a Cartesian coordinates system whereas the Manhattan distance is the sum of the absolute differences, which its name alludes to the grid layout of most streets on the island of Manhattan, which causes the shortest path a car could take between two intersections (see figure 2.6 for a visualization) [69]. In other words in the plane the D_m-distance between points p_1 and p_2 is given by the expression

$$D_m = \left(|x_1 - x_2|^m + |y_1 - y_2|^m\right)^{\frac{1}{m}} \tag{2.6}$$

Thus the Euclidean distance is the L_2-distance and the Manhattan distance is L_1-distance. [68, p. 924] This can be modified to apply to the given RSS vectors

$$D_m = \left(\sum_{i=1}^{n} |RSS_{APi_{Live}} - RSS_{APi_{Ref}}|^m\right)^{\frac{1}{m}} \tag{2.7}$$

where m is depended on the type of distance. Li et al. [69] claims experiments show using Manhattan distance can obtain slightly improved accuracy although in a previous paper Li et al. [70] states that increasing m does not necessarily improve the accuracy of location estimation which while not contradictory claims do infer opposite ideas.

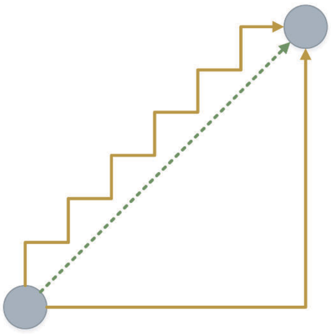


Figure 2.6: Euclidean vs. Manhattan distance: the dotted green line is the direct path (euclidean); the two yellow lines are both Manhattan distances with the same length

Honkavirta et al. [43] uses the Mahalanobis distance M_T in their experiment which measures the distance between a point and a distribution where the general idea is to show how many standard deviations the mean of a point is apart from its underlying distribution. It is defined for an observation $RSS_v = (RSS_1, RSS_2, RSS_3, \ldots, RSS_n)^T$

with observed means of $\mu_v = (\mu_1, \mu_2, \mu_3, \ldots, \mu_n)^T$ as

$$M_T = \sqrt{(RSS - \mu)^T \sum{}^{-1} (RSS - \mu)} \tag{2.8}$$

However they also saw better results using the Manhattan distance instead of the other two described here.

If more than one nearest neighbor is found, i.e. where the D_m-distance is minimal on multiple reference points, either a random position can be picked or to improve on that, the average position of all selected reference points can be calculated. This is known as k nearest neighbour (KNN), where k is the number of minimal distance positions [43].

A further derivation is the k weighted nearest neighbor (KWNN) which uses a weighted average for the estimated location. The inverse of the signal distance defines the weight p which is

$$p = \sum_{i=1}^{k} \frac{1}{L_{mi} + \epsilon} p_i \tag{2.9}$$

where p_i is the position of the k nearest neighbor, ϵ is a small constant to avoid division by zero and L_{mi} is the D_m-distance of position p_i. Also weighting schemes which use the standard deviation σ of the reference measurements exists although they might even degrade the output for initial non-representative RSS readings [70]. To sum these algorithms up, KNN can be seen as a special case of KWNN with equal weights and further Nearest Neighbor (NN) is a special case of KNN where $k = 1$.

A method for improving on moving targets, constrained search-space nearest neighbor optimizes the algorithm in accuracy and performance for this use case. While usually the whole database is queried, with this approach previous locations and time will be taken into consideration and solely reference points that are in a maximum reachable distance will be used as possible candidates [91]. Therefore merely a fraction of the database is needed for the prediction and has the effect of ignoring most likely locations based on Euclidean distance but physically impossible as the next location based on the previous location. On the other hand this algorithm is dependent upon the accuracy of the first measurement as any error there can accumulate over the course of multiple positioning attempts.

Smallest M-Vertex Polygon Smallest M-vertex Polygon Algorithm (SMP) is a technique that forms polygons among selected location candidates and selects the smallest in area using its centroid i.e. the geometric center of the polygon as the estimated location (see figure 2.7). It is thus similar to NN but instead of using the signal domain, it uses the space domain [38, 97].

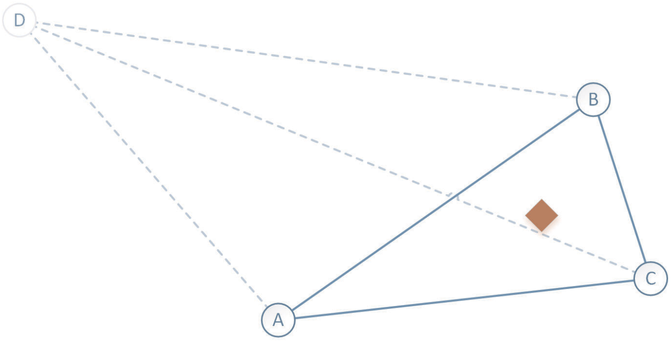


Figure 2.7: The smallest polygon is found between location candidates A,B and C with the diamond shaped centroid as the estimated position of a mobile user

The algorithm uses RSS readings for possible locations from each of the M transmitting base stations. *M-vertex* polygons are formed by choosing at least one candidate from each transmitter [73]. A polygon that which each vertex is from an individual base station is called *distinct-vertex polygon* [86]. The motivation is to allow a fair contribution from all involved transmitters. Then the algorithm selects the polygon with the smallest perimeter i.e. area and returns its centroid. If there is only one coverage from only one base station i.e. a *distinct-vertex polygon* would only consist of one vertex, a random location candidate will be chosen.

Martin et al. shows in his work [77] that SMP fairs approximately on par with KNN with an absolute ≈ 5 % difference in accuracy.

Neural Networks

A neural network can perform complex tasks such as optimization, classification and function approximation, but requires training-data. During the offline phase of fingerprinting, RSS and the corresponding locations are adopted for these purposes. Neural Networks are used as pattern-matching for given RSS vectors of the online phase to return the most likely location. To get the best output internal weights and biases are maintained to minimize a defined error function. Neural networks are usually based on the feedforward Multi-Layer Perceptron (MLP) which is capable of doing a generalization in regions where no training data are available [73]. MLP is composed of a number of interconnected units (neurons) which nonlinearly map the output (locations) from the input (RSS). The coordinates are estimated based on the emitted output of each neuron, which is calculated by the chosen function and weightings. Fang et al. [28] claim a generalized reduction in mean and standard deviation of error of around 20 % over algorithms like KWNN or probabilistic methods. Nerguizian et al. [81] proposed to use Radial Basis Function (RBF) as alternative to MLP because of easier design and faster learning compared to MLP. Also it showed lower location errors during the memorization of the data set and the generalization phase of the network.

Support Vector Machine

Support Vector Machine (SVM) is a fairly new and promising technique for data classification. It is used for statistic analysis and machine learning, and performs very well in many regression and classification applications [112]. SVM already has its utilization in medicine and engineering with positive results. Support vector classification (SVC) of multiple classes and support vector regression (SVR) have been used successfully in location fingerprinting [73]. It can be applied in its regression version to calculate the location of a mobile user in the online phase, and as a classification engine to decide the specific area i.e. a room, in which the user tries to locate himself.

LIBSVM [15] is a library for *Support Vector Machines* written in C++ with different ports in languages like Java, Phyton and Haskell which was used as fingerprinting algorithm in the research of Wu et al. The work [13] of Brunato and Battiti show promising results with this approach applying to fingerprinting achieving slightly better accuracy than with KWNN or Neural Networks (see table 2.1).

Algorithm	Average	75th percentile	95th percentile
SVM	3.04 ± 0.10	3.96	6.09
KWNN	3.06 ± 0.10	3.93	5.79
MLP	3.18 ± 0.11	4.01	6.73
BAY	3.35 ± 0.11	4.39	6.61

Table 2.1: Performance comparison from the work of Brunato and Battiti [13] showing the *Leave-one-out* estimation error distribution

2.1.3 Proximity

Location systems based on proximity use a grid (see figure 2.8) of either antennas [73] or transmitting beacons [25] with a well known position. In the former case the mobile user emits the radio signal whereas in the latter case he receives it. For simplicity it is here reference as *proximity-node* in both cases. The goal is to identify the *proximity-node* that is in closest proximity of the mobile device which serves as the estimated location.

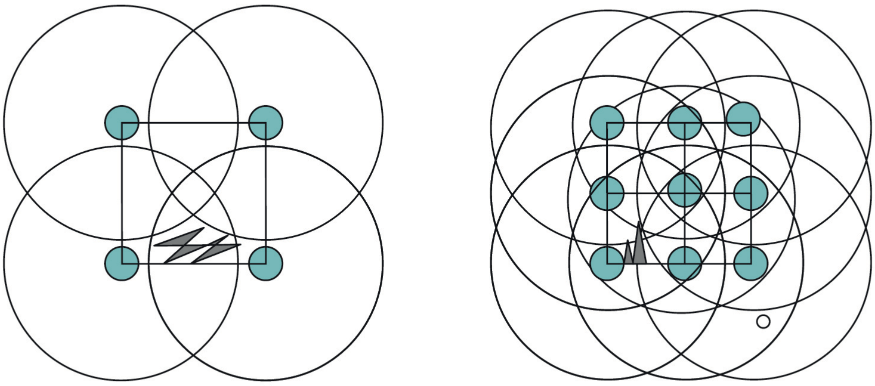


Figure 2.8: The difference in density of transmitters in a proximity based system with overlapping signal radii [14]

To derive the nearest node deterministically, methods that have already been discussed in this work can be applied to RSSI or TOA/TDOA.

Also trilateration (see section 2.1.1) and algorithms like KNN may further enhance the output by averaging the location between *proximity-nodes* if that is required. Apart from that a usual approach is to just select the strongest signal or the shortest arrive time of a radio wave. A proximity-based system is relatively simple to implement over different types of physical media and may incorporate hardware technologies like radio-frequency identification (RFID) [82], Bluetooth (BT) [80] or infrared radiation (IR) [57].

It is therefore a similar concept to fingerprinting with the NN algorithm where the reference points play the same role as the *proximity-nodes* but in a passive manner. In comparison fingerprinting is cheaper and easier to deploy (given a reasonable coverage of radio signals from APs), but may only use the static initial readings of the offline phase whereas proximity grids are always using the current physical measurements and can be adapted to special environments where no coverage of external base stations are available or different accuracies are required in different areas of the target location.

An example of an complex proximity-based location service is discussed in the work of Bulus et al. [14] which is as system that uses beacons as *proximity-nodes* and applies technologies like radio propagation models in combination with RSSI and an positioning algorithm similar to KNN. Kergaard et al. shows in his proposed WLAN-based system TRAX [64] a novel approach of distance measurement by using physical walking distance instead of Euclidean distance for selected cases.

2.1.4 Overview of Related Approaches

One of the earliest and most important research towards indoor positioning with WLAN and RSSI is the in 2000 presented user tracking system RADAR [4] by Bahl and Padmanabhan from *Microsoft Research*. It utilized then common WaveLAN technology, a pre-IEEE 802.11 wireless network which is comparable to 802.11b with the extension that it supports 900 Mhz band in addition to the 2.4 Ghz. The basic architecture consist of inhomogeneous array base stations and one or more mobile client. They state to use *triangulation* in combination with RSSI and propagation models although no real geometric technique is applied. Their "triangulation" method resembles more a fingerprinting approach and is probably just used in lack of a better term at the time. As algorithm Nearest Neighbor (NN) is applied, in the paper described as nearest neighbor in signal space (NNSS), with Euclidean distance although they experimented with Manhattan distance and weighted values. Contrary to most similar systems Bahl and Padmanbahn used the mobile devices to emit radio signals and the fixed base stations to receive them though they claim that the impact of this particular choice to their system is negligible although added complexity is to be expected since multiple nodes need to be time synchronized. As already stated above their basic approach is similar to what is now called *fingerprinting* (see section 2.1.2) with measuring RSS at multiple reference locations in 4 different orientations to account for the human body attenuation. With initial simulated tests they could achieve accuracy of 3 m in 50 % of cases. This could be increased slightly by employing k nearest neighbour (KNN) algorithm (see section 2.1.2) to a median resolution of 2.75 m. The merely small enhancement is explained by the discrepancy of physical positions versus virtual points in signal space. An emphasize is put on the fact that without their orientation specific reference measurements a significant degradation in their service can be experienced to about 4.9 m accuracy. The solution to tracking a moving user is to collect RSSIs of random walks along the test site. Their approach was to use a sliding window of 10 samples to compute the mean signal strength on a continuous basis.

As an alternative to the intricate empirical procedure Bahl and Padmanbahn tested radio propagation models (see section 2.2.3) to extrapolate the data from the empirical method over multiple locations. They choose the Floor Attenuation Factor (FAF) propagation model [100], which is similar to the log-distance path loss model for indoor (see section 2.2.3) with additional empirical factors, because of flexibility in accommodating different building layouts. It was modified in the way to use the wall attenuation but to disregard the path losses through floors which was termed Wall Attenuation Factor (WAF) and described by

$$PL(dB) = PL(d_0) - 10n \log\left(\frac{d}{d_0}\right) - \begin{cases} \mathrm{nW} \times \mathrm{WAF} & \mathrm{nW} < \mathrm{C} \\ \mathrm{C} \times \mathrm{WAF} & \mathrm{nW} \geq \mathrm{C} \end{cases} \tag{2.10}$$

where C is the maximum number of obstructions where the attenuation factor makes a difference and nW is the number of obstruction in the transmitter-receiver (T-R) line and WAF is an empirical factor. To gather the attentions due to walls additional empirical tests were made which concluded in a wall path loss of 3.1 dBm and the maximum number of signal influencing obstructions to be 3. Compared to the empirical method with propagation models a median resolution of 4.3 m was measured clearly worse then the aforementioned but with less effort. In Bahl's and Padmanbahn's continuing work [5] enhancing Radar they addressed issues with mobile user tracking and environment profiling. First they applied a Viterbi algorithm which helps selecting locations that may physically be reached by a moving user and disregard implausible candidates by keeping track of the users location history. A significant improvement of 51 % could be seen in comparison to NN. To address environmental fluctuations the base stations and the knowledge of the physical locations were used to be able to profile the site for different scenarios (e.g. many people versus empty office) with measuring the transmitted signal strengths of each other where they saw considerable improvement in accuracy.

A different approach was presented by Ni et al. with their location-sensing prototype system Landmarc [82]. Unlike Radar, Landmarc (Location Identification based on Dynamic Active RFID Calibration) utilizes RFID as the base technology which uses readers and passive tags. Readers detect tracking tags in a range of 45 m and have additionally a 802.11b interface to communicate with each other. The system employs the idea of having fixed location reference tags to help calibration in order to increase accuracy without having to place more readers. These reference tags are cheaper and therefore can easily be deployed in high volume usually in a grid like array. They are used as a “signal template” for mobile tracking tags to be able to compare them with the actual tag. In this case the radio map consist of the signal informations from the reference tags thus the layout of the reference tags affects the overall achievable resolution. The advantage of this design is that alignment to environmental factors works more reliably because reference information can be dynamically updated through the readers. Pseudo-RSS is derived from the 8 possible ranges provided by RFID front end of the reader when incrementally cycling through them. As algorithm weighted k nearest neighbor (WKNN) is used with Euclidean distance. Using $k = 4$ and weight seen in equation 2.11 provided the least error in most of their experiments.

$$w_j = \frac{\frac{1}{E_i^2}}{\sum_{i=1}^{k} \frac{1}{E_i^2}} \tag{2.11}$$

With a test setup of 4 RF readers, 16 reference tags laid out about every square meter and 8 tracking tags a median accuracy of $1 - 2$ meters could be achieved. With these properties LANDMARC achieves relatively high accuracy wile keeping it low cost. The main drawbacks is a slow position procedure due to workarounds in gathering RSS from RFID and long latencies in sensing of the tag by the location server.

With ARIADNE [56] Ji et al. proposed an automated location determination method that interpolates a RSS radio map with a single signal strength measurement and searches through it with a clustering algorithm. The system has two phases: the map generation and the real time phase where positioning happens. The map generation uses one signal reading and a extensive floor plan. First ray tracing (see section 2.2.1 will be applied in a simpler form without considering diffraction, scattering (see section 2.2), rays under a certain signal strength threshold. Also a simplified multi path model that facilitates propagation modeling is applied. It is defined as

$$PL(dB) = \sum_{i=1}^{N_{r,j}} (P_0 - 20\log_1 0(d_i) - \lambda N_{i,ref} - \alpha N_{i,trans}) \tag{2.12}$$

where $N_{r,j}$ is the total number of rays received at receiver j, P_0 is the reference power value at 1 m, d_i represent the total transmission distance, $N_{i,ref}$ the total number of reflections and $N_{i,trans}$ the total number of walls of the i^{th} ray. λ and α are reflection and transmission coefficient respectively. To estimate these coefficients simulated annealing algorithm, a method used to search for a minimum in a general system, is used which is able to infer the needed values from a single measurement.

System	Technology	Algorithm	Accuracy	Space Dim.	Cost
Microsoft RADAR [4]	WLAN RSSI	KNN	3 to 5 m	2D, 3D	Low
ARIADNE [56]	WLAN RSSI	ray tracing	2 to 4 m	2D, 3D	Low
Horus	WLAN RSSI	Probabilistic	2 m	2D	Low
DIT	WLAN RSSI	SVM, MLP	3 m	2D, 3D	Low
Ekakhau	WLAN RSSI	Probabilistic	1 m	2D	Low
SnapTrack	A-GPS TDOA		5 to 50 m	2D, 3D	Medium
WhereNet	UHF TDOA	Leaste Square	2 to 3 m	2D, 3D	Low
Ubisense	UWB AoA + TDOA	Leaste Square	15 cm	2D, 3D	High
Sappire Dart	UWB TDOA	Leaste Square	30 cm	2D, 3D	High
SmartLOCUS	WLAN RSSI/Ultrasound		2 to 15 cm	2D	High
EIRIS	IR + RSSI		1 m	2D	High
SpotON	RFID RSSI	lateration	depends	2D	Low
LANDMARC [82]	RFID RSSI	WKNN	2 m	2D	Low
TOPAZ	BT RSSI + IR		2 m	2D	Medium
MPS	QDMA	lateration	10 m	2D, 3D	Medium
GPPS	DECT	KNN	7 m	2D	Medium
Robot-based	WLAN RSSI	Bayesian	1.5 m	2D	Medium
MultiLoc	WLAN RSSI	SMP	2.7 m	2D	Medium
TIX	WLAN RSSI	TIX	5.4 m	2D	Medium
PinPOint 3D-ID	UHF RTOF	Bayesian	1 m	2D	Low
GSM fingerprinting	GSM	WKNN	5 m	2D,3D	Medium

Table 2.2: Overview of some indoor positioning-systems with their respective properties [25, 56, 73, 82]

For a efficient map lookup Ji et al. introduced a clustering-based search algorithm. Due to small imprecisions in the generated map errors are to be expected so in order to decide upon multiple valid positions a cluster is formed for every group of location candidates and the largest cluster will be chosen because it has the highest probability to contain the true position. In their experiment they showed an average difference of 3-4 dBm to actual measurements solid values for a relatively low effort though practical measurements are omitted so it is hard to judge how it would fare in real world.

The properties of these systems and others not described in detail can be compared in table 2.2.

2.2 Radio Propagation: Large Scale Models

Radio propagation modeling is the research of predicting radio wave propagation in different surroundings. Traditionally, it is focused on estimating mean received signal strength (RSS) and hence transmission path loss between transmitter and receiver with the auxiliary task of calculating the area of coverage for a transmitter or modeling the distribution of radio signals over different areas. Path loss is the attenuation of radio signal i.e. lost energy through wave propagation.

Transmission paths can vary from simple line-of-sight (LOS) to one that is severely cluttered by obstacles in a building. Unlike wired connections, radio channels are extremely random and do not offer easy prediction nor analysis. Du to multiple variables that add to the propagation behavior like terrain, path, obstacles, atmospheric conditions, frequency and antennas it is not possible to formulate the exact loss for all telecommunication systems in a single mathematical equation. As a result different abstractions exist for indoor and outdoor conditions with some underlying base models discussed in section 2.2.1. Modeling propagation has been historically one of the most difficult parts of mobile radio system designs and typically has statistic-based solutions which are developed from large collections of radio signal readings in different scenarios with diverse frequencies. Hence they do not describe exact behaviors but rather predict likely behavior. [94, 95].

Models that predict mean signal strength for arbitrary T-R separation (i.e. distance) are called *large-scale* propagation models. Contrary *small-scale* or *fading* models address issues over only a short travel distance (i.e. a few wavelengths). Here received signal power may vary by as much as four orders of magnitude when the receiver only moves a fraction of the transmitted wavelength because of small-scale fading and multipath in the mobile radio environment. On the other hand large-scale models do not encounter this vast fluctuation of RSS and are more likely to see a gradual decrease when the receiver moves away [95, p.70].

Applications of propagation models are among others distance estimations, radio coverage design and prediction of signal strength requirements for different scenarios.

The following sections are to familiarize the reader i.a. with basic properties of radio signals and waves thus may contain some basic information about physics. These parts are deliberately kept compact because this is not a discussion about details of physical facts but a summary of data relevant for the next chapters so terms and simple concepts are introduced to the reader. The in-detail explanations are left to the experts in those areas.

Basic Properties of Radio Waves

The physics behind electromagnetic wave propagation is diverse but in general the 3 basic propagation mechanisms are diffraction, reflection and scattering (see figure 2.9) [95, p.78].

Reflection may happen when a wave hits an object that is very large compared to the wavelength. Usually part of the energy is absorbed by the medium of the object and a partial wave is reflected i.e. no energy is lost. If the medium is a perfect conductor all energy is reflected. Thus reflection is depended upon material properties, wave polarization, angle and frequency. For instance radio waves reflect from the surface of the earth or buildings and walls [95, p.78].

Diffraction occurs when the LOS path is obstructed e.g. by a surface that has sharp edges or a slit that is comparable to the size of its wavelength. At these points it appears as new sources of secondary waves. As a result diffraction allows radio waves to propagate around curved surfaces e.g. of the earth. The phenomenon is explained by *Huygen's principle* which states that all points on a wavefront can be considered as point sources for the production of secondary wavelets and these can produce a new wavefront. Like reflection it depends on geometry of objects aswell as phase, amplitude and polarization at the diffraction point [95, p.90].

Scattering is the deviation of radio waves from their main trajectory because its medium contains obstacles that are smaller than their wavelength and the occurrence of these obstacles is high. Scattering also happens on rough surfaces therefore cannot be modeled as reflection, or by irregularities in the channel. As a consequence received signal is often stronger than predicted by reflection and diffraction models alone. In real environments objects like lamp posts and trees tend to scatter waves of cellular towers in all directions. [95, p.100].

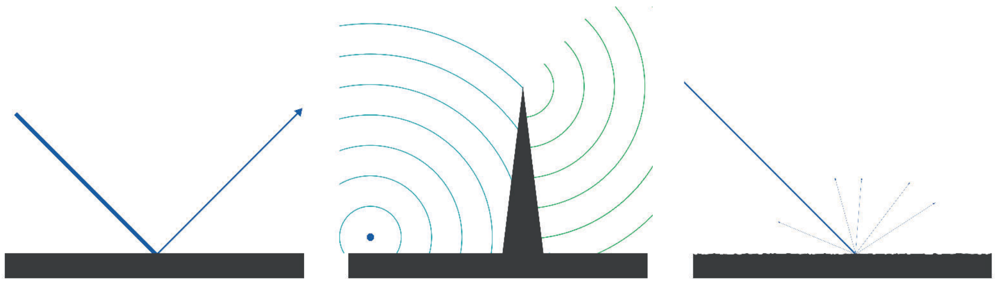


Figure 2.9: Simple examples of reflection, diffraction and scattering

2.2.1 General Models

In the following models are presented that are either very general or can be applied for both indoor and outdoor situations. These models often are used as underlying base for more specific ones. They are usually well researched and accepted as valid assumption in literature however they are often too generic for direct application (without adaptations) and should be more seen as base research of the field.

Free-Space Propagation Model

Free-space path loss (FSPL) is the loss in signal power of a radio wave that would result from a LOS path without any obstruction that could cause diffraction, reflection or scattering (see section 2.2). Formally it is defined as “The loss between two isotropic radiators in free space, expressed as a power ratio” [44] by the IEEE-145-1983 standard for terms of antennas. The model predicts received path loss as a function of the T-R separation raised to some power [95, p.70].

The free-space path loss formula (2.14) is derived from *Friis transmission equation* [32] which describes the received signal strength of a receiver antenna separated by a transmitting antenna by a distance d:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad \Rightarrow \quad \frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi R}\right)^2 \tag{2.13}$$

where P_t is the transmitted signal power, P_r is the received power as function over distance d, G_t and G_r are the antenna gains of the transmitter and receiver respectively, L is a system loss factor and finally λ is the wavelength in meters. We can see that the received power falls off the square of T-R separation [95, p.71].

Therefor the path loss for the free space model without considering antenna gains in decibel (dB) is given by

$$\text{FSPL}(dB) = 10 \log \frac{P_t}{P_r} = -10 \log \left(\frac{\lambda^2}{(4\pi)^2 d^2}\right) \tag{2.14}$$

where the same variable conventions apply as above. This model is only valid in the far-field or Fraunhofer region of the transmitting antenna [95, p.72]. Examples where FSPL applies are satellite communication systems and microwave LOS radio links.

Log-distance Path Loss Model and Shadowing

Theoretical and empirical models indicate that the mean received signal power decreases logarithmically with distance. It is therefore utilized for indoor and outdoor models for different types of sites. This and derived models haven been used extensively in literature [95, p.102]. The log-distance path loss is given for an arbitrary distance d as follows

$$PL(dB) = PL(d_0) + 10n \log\left(\frac{d}{d_0}\right) \tag{2.15}$$

where n is the path loss exponent which indicates the path loss for reference distance d_0 and is depended upon the specific propagation environment - large values mean more attenuation. An example can be observed in table 2.3 and the full list is available in the appendix under table A.1.

Environment	Path Loss Exponent n
Free space	2
Urban cellular	2.7 to 3.5
Obstructed in building	4 to 6

Table 2.3: Some exemplary path loss exponents for the Log-distance path loss model [95, p.104]

The reference path is calculated using the free space path loss formula given by equation 2.14 or through field measurements. This model however does not consider different degrees of clutter in the same distance resulting in vastly different received signal strengths. It has been shown that the path loss at a random location is distributed log-normally about the mean distance-dependent value [9, 21] which is

$$PL(dB) = \overline{PL(d)} + X_\sigma = \overline{PL(d)} + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \tag{2.16}$$

where X_σ is a zero-mean Gaussian distributed random variable with standard deviation σ, both in dB. This describes a random *shadowing* effect which occurs over large sample sizes of signal readings with the same distance d but different levels of clutter in the propagation path. This is also referred to as *log-normal shadowing*. It implies that signal power measurements at specific T-R separations have a normal distribution about the mean of the signal levels of the log-normal distance path loss in dB. It can be used in computer simulations to estimate received power levels in random locations for system designs. Therefor values of n and σ are calculated from measured data using linear regression such that path losses are minimized in a mean square error sense over wide ranges of measurements and distances [95, p.105].

Ray Tracing

In radio propagation ray tracing is the technique of estimating paths of radio waves through space simulating physical mechanics like reflection, diffraction, absorption or scattering (see section 2.2). Waves are modeled as simple rays with a flat wavefront and followed in a defined medium with a specific velocity until they hit an object where the mechanics mentioned above are applied [75]. Increasing the number of rays increase the accuracy but also the computational time. Ray tracing may be simulated in 2D or in 3D with the primary difference being whether wave propagation phenomena at ceilings and floors are considered or not. Nam-Ryul et al. [55] found significant less prediction error with 3D ray tracing in comparison to 2D but not because of consideration of floors/ceilings but because of more counted paths than with the 2D version.

Depending on the simulator ray tracing is able to estimate detailed properties like behavior of electromagnetic waves with different frequencies in different media and the material-types and thickness of obstacles [31] making it a sophisticated choice for predicting radio propagation. The problem with this procedure is though, that with complex scenarios i.e. large number of rooms and walls the computational effort increases exponentially making it unpractical without further optimizing which themselves lower prediction accuracy. Also every environment has to physically be measured and precisely entered in the simulation which might present itself as an endeavor with intricate structures.

Aveneau et al. presented a Java based ray tracing simulator called Radio Propagation Simulator (RaPSor) [3] for narrow band wave indoor propagation (see figure 2.10) targeted to research and educational purposes. It has an extensible build to be able to add geometric primitives, simulation algorithms and new kind of antennas. RaPSor features simulated mechanics like reflection and diffraction of different surfaces and geometries. It was shown to also correctly model ultra wide band propagation in the work of Liѐbe et al. [71]. They saw the simulator obtaining of quick and realistic results and where able to identify the walls, floor, and ceiling with radar of with a ultra wide band pulse.

Generally ray tracing is seen as a sound approach for modeling propagation by literature [27, 31, 55] though without widespread appliances because of its complexities.

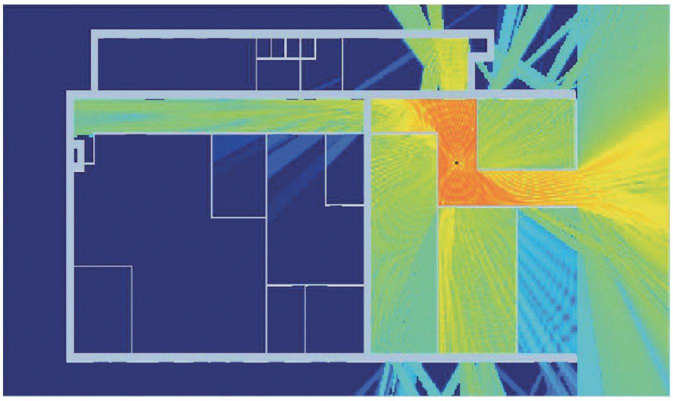


Figure 2.10: An example how RaPSor displays wave propagation [3]

2.2.2 Models for Outdoor Attenuations

In environments where wave propagation is over long range partially obstructed near-line-of-sight paths outdoor models are used. These have to take terrain profile (e.g. mountains), buildings, trees and other obstacles into account. Different approaches exist for predicting mean received signal strengths in these conditions although they greatly vary in complexity and accuracy. Each model aims to cover different types of environments and conditions and are typically only suited for specific scenarios. Most methods are based on empirical measurements and the interpretation and incorporation of this data [95, p.110]. In the following some important models are briefly discussed.

The Okumura model [84] is one of the most widely used models for urban areas. It is applicable over frequencies of 150 Mhz to 1920 Mhz and distance from 1 to 100 km and ideal for cities with many tall buildings. The model is based upon reference functions which give the median attenuation relative to free space which are derived from extensive measurements. It can be expressed as

$$PL_{50}(dB) = L_F + A_{mu}(f, d) - G(h_{te}) - G_{AREA} \tag{2.17}$$

where PL_{50} is the median value of path loss, L_f is the free space path loss, A_{mu} is the median attenuation relative to free space and $G(h_{te})$ is the antenna gain for transmitter and G_{AREA} the antenna gain for the environment [95, p.116]. Okumura's model is considered to have the best accuracy for land cellular mobile systems in crowded metropolitan areas while still being relative simple. The major disadvantage is its slow response to rapid changes in terrain thus it lacks prediction accuracy in rural areas. The usual standard deviation from measured path loss are around 10 dB to 14 dB [95, p.117].

The Hata model [40] is based on the Okumura model that supplies corrections equations for other applications than cluttered urban. It is valid for frequencies of 150 Mhz to 1500 Mhz and employs different corrections for different scenarios which include large cities, suburban and rural area. It does not use any path-specfic corrections like Okumura's model though it compares well in prediction accuracy due to application specific formulas. It works best where the T-R separation is greater than 1 km making it only applicable for very long range communications systems [95, p.119]. The European Co-operative for Scientific and technical research (EURO-COST) with its COST-231 working committed developed an extended version of the Hata model in 1991 to be able to work with frequencies up to 2000 Mhz specifically for 800 Mhz and 1900 Mhz GSM bands [51] which is usually known as Cost231 outdoor model.

More recent research by Green et al. developed a model specifically applicable to LOS WLAN wireless networks, namely the Green-Obaidat Model [36] where a superiority in accuracy to Hata's and Cost231 models is claimed. The path loss formula which is based on Lee's equation for RF propagation [67, p.66] and the FSPL (see section 2.2.1) is expressed as follows

$$PL = 40\log_{10} d + 20\log_{10} f - 20\log_{10} h_t h_r \tag{2.18}$$

where d is the distance of T-R, f the frequency in gigahertz (Ghz) and $h_t h_r$ the antenna heights for the transmitting and receiving end respectively. They showed an offset of only a third dB in their tests compared to the actual measurements. They see the field of applications in mobile ad hoc networks such as vehicles with roof-mounted antennas or wearable computers.

2.2.3 Models for Indoor Attenuations

Modeling indoor propagation can be complicated with the variability in building layout, used materials for walls and floors, interferences with already present radio signal coverages of other technologies and the constant changing environment due to movements of people or opening/closing of doors and windows for instance. In comparison to outdoor models distances are shorter due to the high attenuation in buildings and also because of limited effective radiated power (ERP) of transmitters. Dedicated indoor models may differ therefore considerably although they share the same basic wave propagation mechanics (see section 2.2) as their outdoor pendant [6, 95, 101]

Additional factors like the mounting-height of antennas or presence of obstacles greatly influence the received signal power. Hence prediction accuracy may vary greatly with the combination of site-type and used model. Most indoor models factor different environments with values from reference-tables in to present general solutions. The underlying reference data is usually derived from empirical research by statically interpreting measurements of real sites thus most models are based on empirical studies with combination of general assumptions of electromagnetic wave propagation (see section 2.2.1). Different researches [94, 100] created attenuation tables for various materials and for floor penetration in early years in investigation of indoor propagation which can be reviewed in appendix A.4. Nevertheless indoor models generally hardly ever give a realistic representations of readings measured in real environments but a good estimation that may fit depending on the conditions. Even with sophisticated tools like ray tracing only small derivations from the e.g. the building layout can lead to erroneous predictions.

Indoor propagation models are highly utilized in literature [25], specifically with IEEE 802.11 and similar systems [8, 80] and as such broadly used as solutions for modern positioning systems.

Roughly indoor channels can be categorized in LOS and non-line-of-sight (NLOS) with different degrees of clutter. In the following sections some important ones are described, but mostly in the category of NLOS as this is more relevant for the presented thesis.

Log-distance Path Loss Model for Indoor

A general-site model i.e. applicable for most indoor environments the log-distance path loss model already discussed in 2.2.1 is a modified power law with a log-normal variability, similar to log-normal shadowing. It is given by the equation

$$PL(dB) = PL(d_0) + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \tag{2.19}$$

where n is a coefficient depending upon the building type (see tab. 2.3), X_σ is a normal random variable in dB with zero mean and standard deviation σ dB reflecting the attenuation caused by flat-fading. In case of no fading the value is 0, d is the path length, d_0 a reference distance (usually 1 km) and $PL(d_0)$ is the path loss at the reference distance d_0 [95, p.127].

Chrysikos et al. [18] has shown that the shadowing effect described by equation 2.19 can be also measured under 2.4 Ghz IEEE 802.11 wireless networks. They showed that it coincides with the measured values in complex indoor topologies containing same and multi level floor setups.

Ericsson Multiple Breakpoint Model

Published in 1989 for Digital Enhanced Cordless Telecommunications (DECT) telephone systems by Akerberg [1] it was tested for 900 Mhz but is also intended to be used for the 1.6 Ghz band. Their propagation model was derived from radio signal readings in their office building where transmitters were located in the middle corridors and receivers in rooms along these corridors and in different stories with a added Rayleigh distribution. Isolation between floor was set to 10 dB and is limited to 3 stories. The model has four breakpoints and assumes a minimal attenuation of 30 dB at 1 m. It provides a deterministic limit of the range of path loss at particular distances rather than assuming log-normal shadowing [95, p.128]. The results can be seen in figure 2.11 as a function plot of path loss.

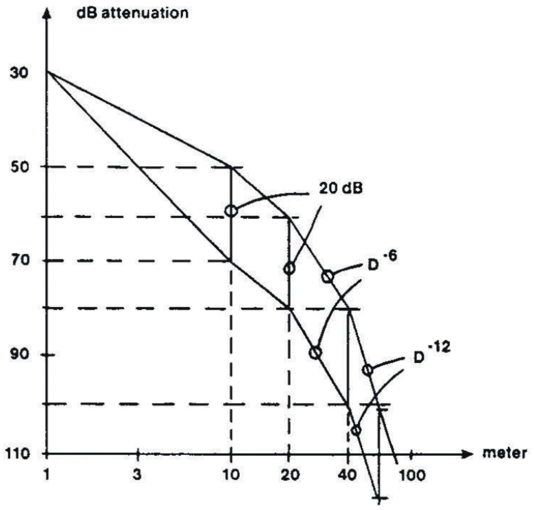


Figure 2.11: The resulting Graph of the Ericsson Multiple Breakpoint Model [1]

ITU model for indoor attenuation

Developed by Keenan and Motley [62, 63] in 1990 this model takes site-specific effects of greater transmission loss into consideration by factoring walls, floors and ceilings in; it is therefore also categorized as multi-wall model [6, p.314]. The model was also adopted by the International Telecommunication Union (ITU) in their recommendation for radio propagation in ITU-R P.1238 [92] currently in its seventh edition (2012). It is applicable for frequencies of 900 Mhz to 100 Ghz but has its best performance at microwave wavelength specifically verified in experiments for frequencies used by IEEE 802.11 wireless networks i.e. 2.4 Ghz and 5 Ghz [19].

$$PL = 20\log_{10} f + N\log_{10} d + P_f(n) - 28 \tag{2.20}$$

where f is the frequency in megahertz (Mhz), d the T-R distance in meters, N is a power loss coefficient provided by reference tables (see tab. 2.4), $P_f(n)$ is the floor loss penetration factor (see tab. 2.5).

Frequency	Residential	Office
2.4 Ghz	28	30
5.2 Ghz	30	31
5.8 Ghz		24

Table 2.4: Excerpt of the empirical variable N for the power loss coefficients for different conditions [92, p.4]. The full list can be found in the appendix A.2.

Frequency	Type	Residential	Office
2.4 Ghz	Apartment	10	14
2.4 Ghz	House	5	
5.2 Ghz	Apartment	13	16 (1 floor)
5.2 Ghz	House	7	

Table 2.5: Excerpt of the empirical values for the floor penetration loss factors [92, p.5]. The full list can be found in the appendix A.3.

This model is described as is the most practical and easy to implement among semi-empirical indoor propagation models by Chrysikos et al. [18] though it is also stated to overestimate propagation losses in cases where direct paths between links is obstructed by several floor/walls and indirect paths become more significant although not accounted for in the model [6, p.315].

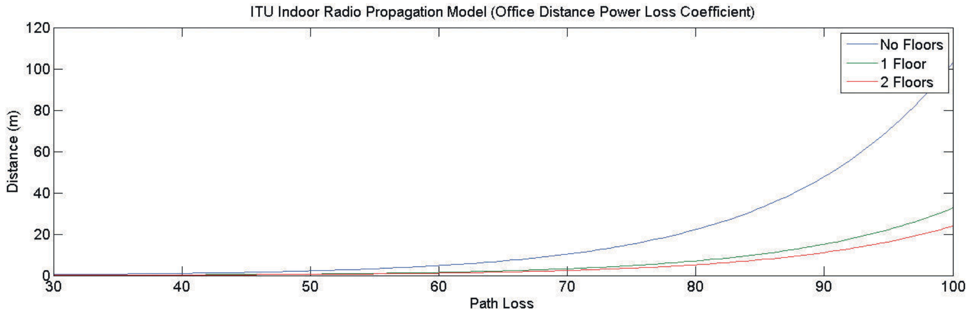


Figure 2.12: ITU model path loss plotted as a function of distance for office types at 2 Ghz

A correction [23, p.176] for the issue of excessive path loss through multiple floors is included in one of the narrow band models adopted by the COST-231 committee where the floor loss coefficient P is provided as

$$P_f(\text{effective}) = P_f^{\frac{P_f+2}{P_f+1}-b} \tag{2.21}$$

where b is an empirical parameter suggested as 0.48. This is based on the work of Karlsson et al. [61] where a value of 0.5 is suggested for b with slightly different fraction term as the power of P_f i.e. $\frac{P_f+5}{P_f+3}$ which itself is based on the work of Bertoni et al. [10] from 1994.

2.3 Analysis of RSSI Method

WLANs, also called Wi-Fi based on the IEEE 802.11 standard became very cheap in recent years making them widely available. This is the reason why RSS based location services is a popular solution since it hardly requires new hardware. To better understand the properties and behavior of RSSI this section provides some background information. It covers the used unit, discusses some details of WLAN itself and then analyzes statistical properties of RSS in IEEE 802.11.

The terms received signal strength (RSS) and received signal strength indicator (RSSI) seem to be used interchangeably, though they are not synonyms. RSSI describes a metric of received signal strength while on the other hand RSS stands for the actual value of measured RSSI. This is just to clarify and prevent misunderstands for the upcoming sections.

All the discussed topics here relate to RSSI in the IEEE 802.11 standard, although most concepts also apply to other types of wireless networks (e.g. Bluetooth). The reason being, that this is the most relevant to the presented thesis and also the most thoroughly researched in science publications in addition to the IEEE 802.11 standard itself. So actually the emphasis in this part is RSSI specifically in IEEE 802.11.

As with the previous section this part contains basic information about physical properties although again in a very brief way. It is important to know certain reference values and circumstances to be able to better interpret results and experiments conducted in this thesis.

2.3.1 Background

RSSI is an indication of the power level being received by the antenna. Thus the higher the number, the higher the power level, the stronger the received signal. There is actually no standardized relationship of any physical unit to RSSI measurements as defined in the IEEE standard. 802.11 does not define any correlation between RSS value and power level [49]. Vendors provide their own accuracy, granularity, and range for the actual power and these may differ significantly (see also section 2.3) [76]. The typically used unit is decibel-milliwatts (dBm) which is a ratio between the logarithmic scale dB and milliwatt (mW) as a unit of transmitting power where 1 mW is 0 dBm (see eq. 2.22) [99].

$$P_{dBm} = \log P_{mW} \times 10 \tag{2.22}$$

It is a convenient measure for this purpose as it can express values of different magnitudes in an easily readable manner (see table 2.6) e.g. 100 dBm are equivalent to 1×10^{10} mW whereas -10 dBm equals 1×10^{-4} mW. As a rule of thumb one can remember that the power in mW approximately doubles or gets cut in half every 3 dBm that is added or subtracted respectively which might be unintuitive at first when one encounters a RSS value like -47 dBm ($\approx 2 \times 10^{-5}$ mW) which e.g. is only half the signal strength in mW of -44 dBm ($\approx 4 \times 10^{-5}$ mW). As a reference in an IEEE 802.11 wireless network usual RSS values are between -30 dBm and -90 dBm [76].

These values are usually converted to a more comprehensible linear unit namely signal strength percentage or signal quality on a user interface level. Unfortunately there is no standardized conversation algorithm so every software vendor and chipset makers may implement their own method. For example the *Windows* operating system defines this

Level	Power	Description
30 dBm	1 000 mW (1 W)	Max. EiRP of 802.11 according to FCC [29]
20 dBm	100 mW	legal limit of EiRP in 802.11 [30]
15 dBm	32.623 mW	EiRP of average WLAN device [30]
10 dBm	10 mW	
3 dBm	2.0 mW	
1 dBm	1.3 mW	
0 dBm	1.0 mW	Bluetooth standard (Class 3) [11, p.8]
−1 dBm	0.794 mW	
−3 dBm	0.501 mW	
−10 dBm	0.000 1 mW (100 μW)	Maximum RSS in WLAN 802.11 [49, p. 1580]
−30 dBm	0.000 001 mW (1 μW)	
−50 dBm	0.000 000 01 mW (10 nW)	
−80 dBm	0.000 000 000 1 mW	
−100 dBm	0.000 000 000 001 mW	Minimum RSS in WLAN 802.11 [49, p. 354]

Table 2.6: An overview of power levels in the logarithmic scale dBm in relation to power in mW. Values are rounded to the third fraction

with a minimum and maximum value for 0 % and 100 % respectively while linearly interpolating the values in between [79]. On the other hand the network-device manufacture *Cisco* uses a lookup table and its competitor *Atheros* a simple formula [7]. Examples of the different scales can be seen in table 2.7.

RSSI is basically depended upon the ERP, or more specifically Equivalent isotropically radiated power (EiRP) [105, p.111] which is composed of the tranmission power output (TPO), internal cables and antenna gain of the transmitter as well as sensitivity and antenna gain of the receiver [105, p.110]. In indoor environment especially the free space path loss plays an important role because it is the most relevant value that is needed to be estimated for a distance calculation (see section 2.2). This can be visualized as the

RSS Percentage	Microsoft	Cisco	Atheros
100 %	−50 dBm	−10 dBm	−35 dBm
75 %	−62.5 dBm	−33 dBm	−50 dBm
50 %	−75 dBm	−58 dBm	−65 dBm
25 %	−87.5 dBm	−87 dBm	−80 dBm
0 %	−100 dBm	−113 dBm	−95 dBm

Table 2.7: Comparison of signal strength percentage interpretations of different vendors

following simple summation seen in equation 2.23 [36]

$$\text{RSS}_{\text{dBm}} = \overbrace{\underbrace{\text{TPO}_T}_{\text{dBm}} - \underbrace{\text{CL}_T}_{\text{dB}} + \underbrace{\text{AG}_T}_{\text{dBi}}}^{\text{EiRP of Transmitter}} - \overbrace{\underbrace{\text{Pathloss}}_{\text{dB}}}^{\text{Free Space}} - \overbrace{\underbrace{\text{CL}_R}_{\text{dB}} + \underbrace{\text{AG}_R}_{\text{dBi}}}^{\text{Receiver}} \tag{2.23}$$

where TPO is the emitted power of the radio chip in dBm, CL is cable-loss which is like path-loss measured in dB whereas antenna gains (AG) use the unit Decibel-isotropic (dBi). Typical TPO levels of a WLAN 802.11 devices are usually between 15 dBm (32 mW) and 20 dBm (100 mW) [30, p.22] and may be dynamically set by the user. Reference values are around 2-5 dBi antenna gain for a conventional AP [74, p.15] and 0.5-2 dB for cable loss per meter [88].

In this equation a perfect isotropic antenna is assumed which theoretically radiates the signal uniformly in a sphere shape in all directions (which is only a hypothetical model and not physically possible) [105, p.50]. This is the reason dBi is used in these equations. However by default most network devices are assembled with an omni-directional class antenna (e.g. dipole) which often have a radiation pattern of a torus (see figure 2.13) [105, p.37] along a given plane with the radiated power decreasing with elevation angle above or below the plane, dropping to zero on the antenna's axis. This is the most practical approach since it does not require the antenna to be specifically directed to the receiving parts of the wireless network. Contrary a directional (e.g. parabolic) antenna focuses the signal power into a narrow cone-shaped beam which results in longer ranges, but smaller areal coverage [105, p.17]. These kind of radiators are utilized to bypass long distances to a specific node in the network and therefore to extend its reach while usually having a much higher gain value where 24+ dBi are not uncommon [93]. This is also utilized to supply large areas in developed countries with internet access as shown in the work of Raman et al. who studied a project in India. Many other types of antennas do exists although not relevant for the topic discussed here thus omitted in this work for brevity.

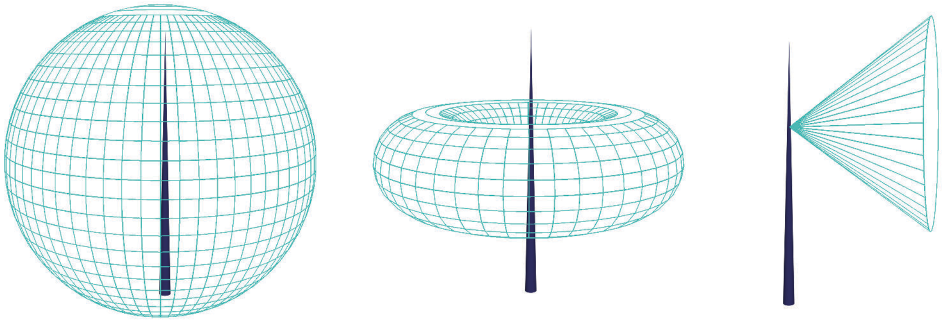


Figure 2.13: From left to right: isotropic pattern [105, p.18], omni-directional pattern [105, p.50] and directional pattern

Wireless Networks IEEE 802.11

IEEE 802.11 as the widest used standard for wireless networks has seen many amendments since its introduction in 1997 that increased throughput, frequency usage, security and antenna utilization among others. The 1999 approved 802.11b was the first widely adopted by routers, APs and clients. It featured a then vast gross data-rate of 11 Mbit/s but lacked security with only supporting weak encryption. 802.11a tried to defuse the crowded 2.4 GHz band by using the 5 GHz instead though sacrificing signal range. A couple of years later in 2003 802.11g enhanced the bandwidth to 54 Mbit/s [46] and in the same year the the Advanced Encryption Standard (AES)-based security protocol Wi-Fi Protected Access (WPA) emerged as *draft 802.11i* preceding the current state-of-the-art Wi-Fi Protected Access II (WPA2) in the final *802.11i* a year later [47]. In 2009 802.11n enhanced the standard by additionally extending to the 5 Ghz band and allowing multiple antennas through multiple-input and multiple-output (MIMO) for a boosted data rate up to 600 Mbit/s [48]. One of the most recent additions is provided by 802.11ac in 2013 which i.a. broadened the channel width to 80/160 Mhz in the 5 Ghz band for further increasing throughput [50].

To summarize wireless networks implementing IEEE 802.11 work with four distinct frequency ranges: 2.4 GHz, 3.6 GHz, 4.9 GHz, 5 GHz, and 5.9 GHz bands [49], where 2.4 GHz is the most widespread used because its one of the first introduced [22] and 5 GHz gaining compatibility on most devices.

The 802.11 work-group documents 14 channels in the license-free 2.4 GHz Industry, Scientific and Medical (ISM) band from 2412 Mhz to 2484 Mhz. According to national regulations not all of these channels are allowed for usage e.g. in Europe only 13 channels may be used while in North America the channel 12 and 13 are only allowed under low powered conditions [49, p.1572]. Because of the limited frequency range channels will overlap with adjacent ones and interfere with each other (see figure 2.14). In 802.11b only 4 non-overlapping channels do exists while in 802.11g there are only 3 and finally in 802.11n there is only a single one due to the wider 40 Mhz channel width [48, p.224] versus the more narrow 20/22 Mhz [49, p.1576] in the other two standards.

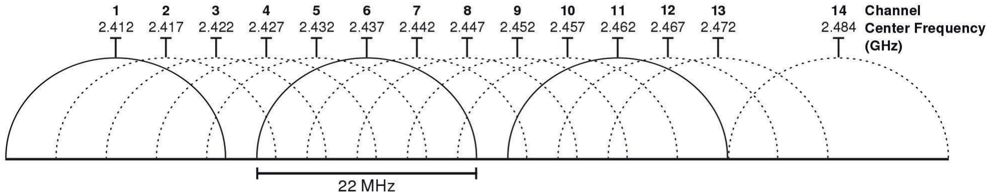


Figure 2.14: Overlapping channels of IEEE 802.11b with a channel width of 22 Mhz. Only channels 1, 6 , 11 and 14 are non-overlapping [33]

In the 5 GHz ISM band, which is advantageous to the crowded 2.4 GHz frequency (Bluetooth, microwaves, etc) because of less interference but suffers of lower ranges, there are up to 200 channel [48, p.313] in 5 Mhz intervals but only a few are approved depending on the regulations of the respective region. For example in Europe which is regulated by the European Telecommunications Standards Institute (ETSI) the channels 36 (5180 Mhz), 40 (5200 Mhz) and 44 (5220 Mhz) i.a. are open to use [26]. Because of the wider range it is possible to provide more non-overlapping channels and also support significant wider ones e.g. 160 Mhz in 802.11ac.

In infrastructure mode an AP with all associated clients is called basic service set (BSS) [49, p.7]. The AP acts as a master and as such has different information to broadcast to identify itself. The service set identifier (SSID) is a set of 32 octets that is treated as encoded with UTF-8 [49, p.478] (with exceptions) and usually is a human readable identifier for wireless networks (but not its main purpose). An AP may propagate multiple SSIDs for different frequencies or virtual local area networks (VLANs). To uniquely identify a BSS and its AP basic service set identifier (BSSID) is utilized [49, p.108]. This is the 48 bit media access control (MAC) address of the radio chipset typically represented as 6 groups of two hexadecimal digits split by a colon “`:`” or hyphen “`-`” (e.g. `01:23:45:67:89:ab` or `01-23-45-67-89-ab`) as defined by IEEE 802 [45, p.25]. It is also possible that different BSSs are interconnected and share the same SSID and security settings which is then called a extended service set (ESS) [49, p.47] and thus the SSID is referred to as extended service set identifier (ESSID). Another type of network is the *ad hoc network* which lacks a dedicated master node and directly connects clients, formally also called independent BSS (IBSS).

SSID	UPC013652
BSSID	`00:24:D1:83:1F:3A`
Channel	6
RSSI	−83 dBm
Encryption	mixed WPA/WPA2 PSK (TKIP, CCMP)

Table 2.8: An extract of the provided information for a single AP of a WLAN survey by the tool `iwinfo`, a command-line front-end developed for the embedded Linux distribution OpenWRT [85]

2.3.2 Statistical properties of RSS in IEEE 802.11

In order to identify if RSSI is a consistent measure usable as a reliable indicator its statistical properties of measurement have to be evaluated. As a basic behavior it can be said that multipath fading causes the received signal to fluctuate around a mean value at particular locations [58]. On the physical level the average RSS is believed to be log-normally distributed according to popular large-scale-fading models [103]. However there is correctly no consensus regarding the RSS distribution measured at the software level by the wireless network interface controller (NIC) and standard deviations of signal strengths are not understood well [58].

Distribution

To evaluate the distribution empirical studies have to be done. Considering the complexity of such experiments and completely different setups it is not surprising that various opinions exist in literature. For instance with IEEE 802.11 Small et al. [104] interpreted their readings as normal distribution since they registered that the values of mean, median and mode are very similar or the same which is a characterization of said distribution. This can be examined in figure 2.15. Other researchers though disagree. Kaemarungsi et al. found that normal distribution may only apply in certain situation but not in others which is also confirmed by Chen et al. [17]. They actually recognized most of their measurements as being left-skewed approximated Gaussian distributions and attributed that to their test setup being mostly NLOS in far proximity of the transmitter. In their later work [58] they further found out that no observable relationship between the RSS variations transmitted by different APs exist and therefore assumed independence.

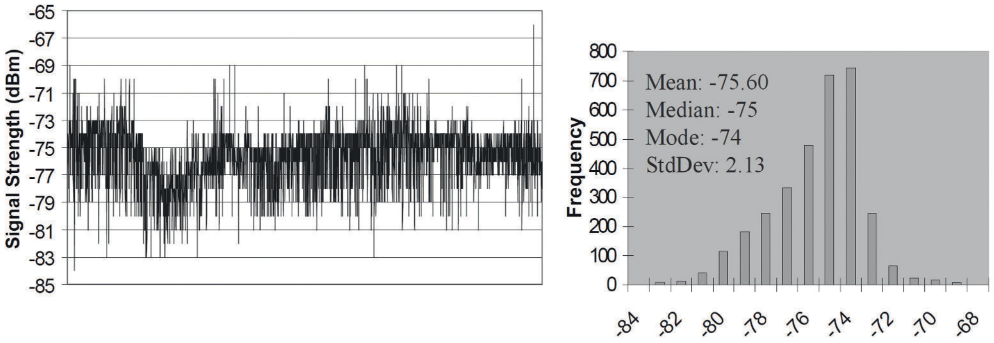


Figure 2.15: Plotted RSS readings of a single AP for 5 s interval over about five hours and the underlying frequency distribution [104]

Ladd et al. [66] observed that distributions were asymmetric and had multiple modes. Usually there was a dominant mode which often differed from the mean. The conclusion was that distributions were essentially non-Gaussian and consequently they used the observed distribution directly instead of reducing it to a model and using average values.

As a result of the complexity of radio propagation, the distribution of RSS is difficult to model and fit to well-known distributions. Parameswaran et al. [87] even concluded that RSSI can not be used as a reliable metric in localization algorithms because of its inconsistent behavior, its complex attenuation and the gain of error rate with increased distances and unreliable results in extreme ranges.

RSS as indicator for Distance

The topic of using RSS as distance indicator is specifically relevant if geometric positioning techniques are applied in combination with RSS (see section 2.1.1). As stated above RSSI is an inconsistent measure although for distance estimation additional challenges present itself. As established by different studies [87,104] signal strength readings become more unreliable the further the distance. It has therefore be taken care to take this into consideration when interpreting power readings. For instance a higher degree of confidence can be laid on stronger signals which can bee seen in figure 2.16. However in the work of Kaemarungsi and Krishnamurthy [58] the contrary is observed with higher power signals that suffering from increased standard deviation than lower ones. As a different approach to improving on the fluctuating readings of RSSI Teuber et al. [107] proposes improvements by incorporation fuzzy logic to filter out invalid measurements. They saw clear improvements in accuracy but also admitted that their system needs to be adapted for the actual environment.

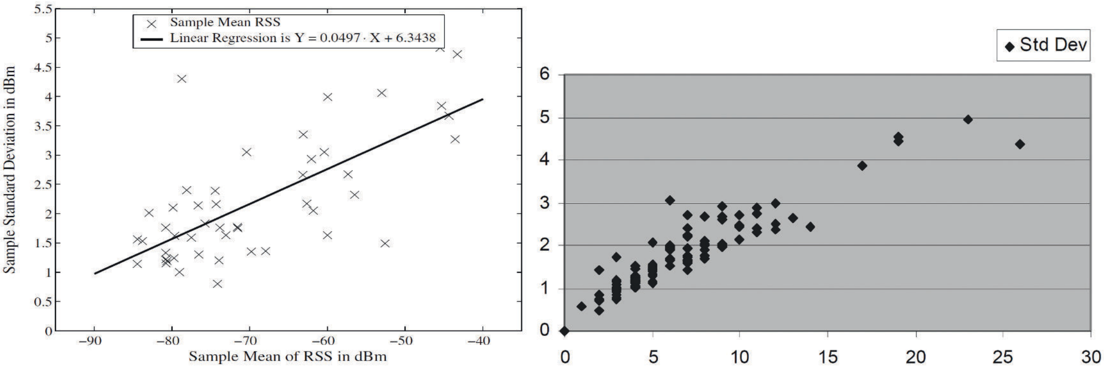


Figure 2.16: The relationship between average RSS and its standard deviation: on the left the observation of Kaemarungsi et al. [58] is showed and on the right measurements of Small et al. [104]. The x-axis of the right plot is a distance measurement whereas the y-axis represents the standard deviation value.

2.3.3 Additional Properties

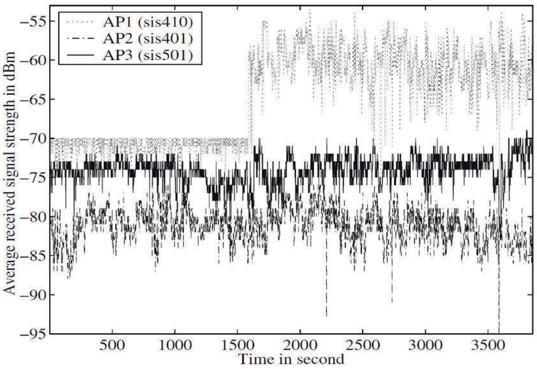
A non exhaustive list of additional properties and challenges when working with RSS follows.

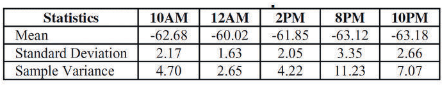
Interference of Multiple APs

It is assumed by the way IEEE 802.11 MAC works that multiple concurrent APs do not interfere with each other since transmissions are either not heard or deferred. This can be witnessed in experiments by Kaemarungsi and Krishnamurthy [58] where RSSs of three different APs were measured over multiple days with uncorrelated behavior (see figure 2.17).

Time dependency of RSS

An important question is if RSS readings are constant over prolonged time. This is called stationary in statistics i.e. if mean, median, etc. are all constant. Current research [58,104] does see this property in experiments over hours and days concluding that signal strength





Statistics	**10AM**	**12AM**	**2PM**	**8PM**	**10PM**
Mean	-62.68	-60.02	-61.85	-63.12	-63.18
Standard Deviation	2.17	1.63	2.05	3.35	2.66
Sample Variance	4.70	2.65	4.22	11.23	7.07

Figure 2.17: RSSI of 3 different APss over the course of multiple days [58]

does not vary dramatically at different times of the day. These fluctuations are mostly attributed due to the presence of people which interfere with the signals.

Impact of Human Body

With personal based indoor positioning systems it can be assumed that at least one part of the system is carried by the user. This fact can play an important role for measuring signal strengths since the frequencies used interferer with the human body and thus get absorbed (in the simplest case) to a certain degree. When applying fingerprinting techniques (see section 2.1.2) this does of course not only affect the online live positioning phase but also the offline reference measurement phase.

Experiments by Zhang et al. [115] (see figure 2.18) show that this is clearly an observable effect. They saw a up to 18 dB difference in RSS readings when a tester holds a WLAN-enabled smartphone in front of him and rotates with a speed of 6° per second for a full 360° with an AP in about 2 m LOS. In their tests they did not observe any measurable difference when differing the mounting-height of the AP and also ruled out receiver antenna placement i.e. smartphone as the source for this effect by shifting its orientation without an immediate presence of a person. The tests were confirmed by repeating it with different hardware and persons of varying statues. Similar conclusions are also found by other researchers [99, 114]

The problem is that with microwave-type radio signals of IEEE 802.11 and with 2.4 Ghz specifically that it is the resonance frequency of water and therefore for human bodies (≈ 65 % water) which means that they absorb RF signals [114]

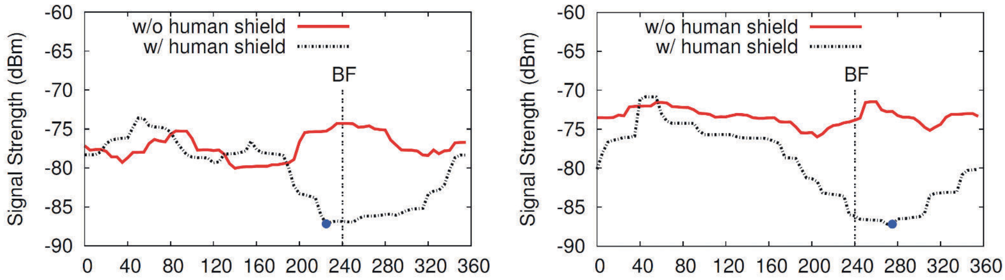


Figure 2.18: RSS readings plotted over orientation in degrees with two different devices. “BF” means that the user faces the AP i.e. the transmitter with his back [115]

2.4 Analysis

It can be summarized that RSS with WLAN is a suboptimal indicator for distance and therefore positioning because of its semi-predictable nature. Nevertheless as can be compared in the previous subsection 2.1.4 “Overview of Related Approaches” it is often used as such, with approaches that try to circumvent the complexity of indoor radio wave propagation or model its behavior with complex methods like ray tracing. As a middle ground, propagation models may be chosen which have shown to generally work in various conditions, provided the appropriate model is picked. They, however, still only represent a rough estimation that does not take subtleties like attenuation near walls versus in the middle of a room into account. As for positioning techniques, scene analysis seems to be the best compromise between effort, complexity and accuracy. Compared to it, proximity may achieve considerably higher accuracy, but with the added cost and effort to provide a narrow grid. Geometric methods like trilateration require an exact distance estimation for the most part to provide acceptable results. In combination with RSS and IEEE 802.11 this will limit its output to only rough estimations and will generally not be a satisfying option as a direct method for positioning. All these findings directly influenced the design of the proposed system which will be detailed in the following chapter.

CHAPTER 3

Design and Implementation

This chapter describes the design and implementation details of the proposed indoor positioning system called *Site Evaluating Active Sensor Network Indoor Positioning System* (SEASNIPS) . It is a standalone service, offering one complete solution, from measurement to actual user positioning. All components and procedures that have been used are discussed and where not obvious, justification for specific design choices is provided.

The design is based upon predictions resulting from analysis of the accumulated research presented in chapter 2 “State of the Art”. The appropriated positioning techniques will be used at different stages in the system according to the anticipated benefit. Since in the architecture active elements are used that may continuously measure each other, a graph can be generated that represents the underlying distance matrix. For this, a simple trilateration algorithm seems appropriate. This technique is considered to not be very accurate in combination with received signal strength (RSS), but it is predicted to generate usable results in most cases where it is expected that the outcome will be manually adjusted regardless. To provide a reasonable conversion from signal strength to distance, radio propagation models where selected as a good compromise between complexity, accuracy and practicability. Although in this context the more accurate ray tracing method could appear appropriate, the requirement of providing a correct three dimensional mapping of the target area renders this solution rather impractical in large-scale applications. Additionally it requires enormous computational power. Based on findings presented in subsection 2.2.3 “Models for Indoor Attenuations” and through testing, a radio propagation model was chosen which fits the requirements of this system.

It is anticipated that the fluctuating nature of RSS can be partially compensated by using measurements from active elements distributed throughout the site, in combination, which will be able to adapt the propagation model to generate an average representation of its attenuation. With the model already customized the logical choice was to use interpolated signal propagation to generate a map that represents the theoretical RSS values at every location in the target area. The map is very similar to those provided in the fingerprinting approach so the same positioning algorithms could be applied. After analyzing the various available options it was decided to use a variation of the nearest neighbor algorithm (see section 2.1.2) as it appeared to provide reasonable results with low complexity.

To keep the implementation within the boundaries of limitations a thesis imposes the following restrictions are set:

- The target area is simplified to a 2D planar. This significantly reduces computational cost and complexity of the system. It is also rather questionable whether it would be effective to work in the third dimension, particularly given the considerable additional effort this implies compared to the uncertainty regarding the resulting increase of accuracy in the system.
- Only single floor mapping is supported. That means that every node used for the generated map has to be in the same floor to avoid incorrect readings due to the conditions explained in the point above. The system is designed to manage an arbitrary amount of maps so as to allow for the simultaneous use of multiple floors for positioning without the deficiency of shared or overlapping nodes. This is also to avoid the complication of complex floor attenuation in buildings.
- The problem of interference of human bodies with radio waves is ignored (see section 2.3.3 “Impact of Human Body”). All measurements are performed without a person in near vicinity. This, of course, is not representative for real life application, however the main aim of this work is to illustrate the proposed system’s general performance.
- Support is limited to user *positioning* and user *tracking*, referring to moving users, is deliberately excluded.
- No active client is implemented. In its current state the server handles and displays the positioning of clients. Currently clients only act as passive devices similar to active nodes (see section 3.1.2 “Positioning Client”).

All of these arguments are discussed in detail in chapter 5 “Conclusion and Outlook”.

3.1 Design

For enhanced consumption all phases and components of the system, presented in the design section, are complemented with instructive diagrams. Details on specific approaches are then duly discussed in the following section concerning the implementation.

3.1.1 Architecture

Seasnips is an IEEE 802.11 with received signal strength indicator (RSSI) based indoor positioning system. Seasnips consists of a service control server (SCS), an array of active nodes (ANs), i.e. Wireless Local Area Network (WLAN) access points (APs), zero or more extended nodes (ENs) and tracking clients. Active nodes are under the control of the system’s administration and contain software packages that facilitate the device’s hardware features and are responsible for network communication with the server. Extended nodes are regular APs usually present in urban environments where their placement is not under the control of the service provider. The server handles continuous RSS measurements, manual as well as automatic site-adaptations and the map generation. Its positioning module handles the live phase where clients try to use the service to locate themselves. Clients are arbitrary IEEE 802.11 WLAN enabled devices that are capable of accessing the server’s service application programming interfaces (APIs). A simplified representation of the system described above can be examined in figure 3.1.

First the system has to be set up with reasonable placement of ANs, i.e. it should cover all of the target area and be visible to at least three other nodes. When the SCS is initialized and connected to the nodes RSS measuring can begin. After a sufficient amount of cycles have been run, to derive stable statistical properties from the sample set, the next phase, which will be site adjustment, can be started. Hereby first all active nodes need to be normalized for their differing RSS at same distances. As a next step the used propagation model has to be adjusted for all known distances to better suit the current site properties. After that the map generation phase is initialized by approximating the active nodes’ positions via trilateration. Manual correction for the x- and y-axis has to be performed. The position of extended nodes can be derived again with trilateration. Now that all properties are set, the system can calculate the radio propagation of all nodes based on the previously adjusted model. With this information available, a radio map can now be

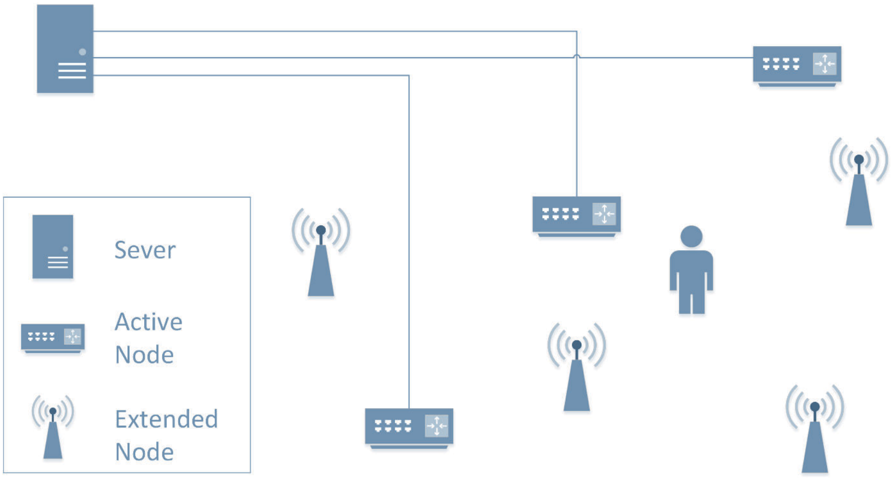


Figure 3.1: A simple outline of the systems architecture

generated for the live positioning phase. The general method used here is scene analysis with Nearest Neighbor (NN) algorithm. The intention is hereby to repeat all steps from measurement to map generation continuously to steadily adapt to environmental changes. A diagram of these phases be observed in figure 3.2.

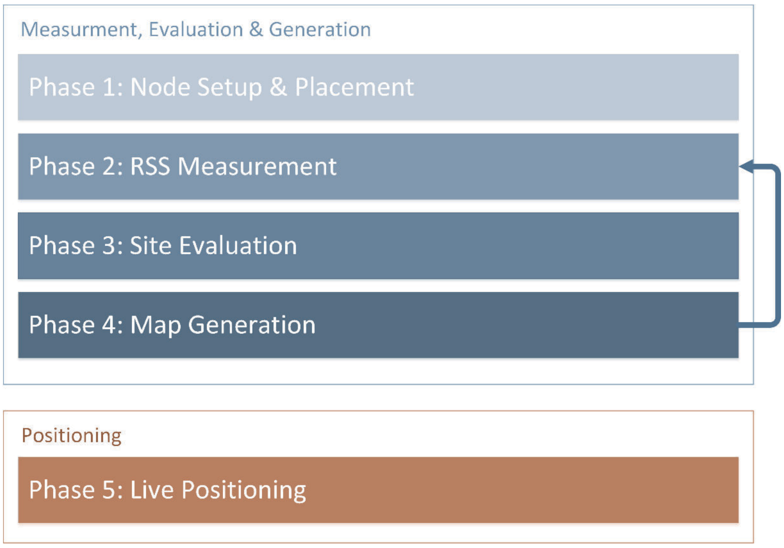

Measurment, Evaluation & Generation
Phase 1: Node Setup & Placement
Phase 2: RSS Measurement
Phase 3: Site Evaluation
Phase 4: Map Generation
Positioning
Phase 5: Live Positioning

Figure 3.2: A simple outline of the systems phases

Goals and Aim

It is important to understand the goals of the proposed system. Because of the used technology it is not expected to achieve sub-meter accuracies as with systems with specialized hardware (e.g. LANDMARC [82] by Ni et al.), thus highest achievable accuracy can not be the main goal of this exercise. Therefore the aimed resolution is on the sub-room level. The main emphasizes is to simplify the setup and make it easy to measure large areas while integrating already deployed hardware. As such the main properties are as follows:

Scalability Because the system is maintained on a single server after the initial setup it is very easy to expand and adapt.

Cost WLAN is broadly available and cheap. At the moment only open or free software is used so no license fees are to be paid. Also hardware requirements for the server are very low by today's standard.

Robustness Due to constant re-measuring and resulting re-adjusting of the system, it is possible to adapt within hours even to rapid environmental changes in only hours. This makes the service very robust to variance and fluctuation.

This leads to the intended field of application in sectors where large areas need to be maintained from a central unit for low hardware maintenance and only limited requirements for accuracy. This would fit service types like personal navigation in airports or train stations or context-aware services in shopping malls.

Design Decisions

After reading through the architecture description it may seem odd that some techniques are used which are described in chapter 2 "State of the Art" as "not optimal" to be modest. Therefore to help the reader understand the reasons for resorting to certain methods the following list of important design decisions are rationalized. Admittedly some decisions are made based upon lacking suitable alternatives, though most of them present satisfying results nonetheless.

IEEE 802.11 as network The reason to use this technology is simple and twofold: cost and availability. Devices are cheap and thanks to open software availability, easily customizable. Most buildings already present a decent coverage of WLAN and APs. This is important as it illustrates the available opportunities for reducing the necessity for installing additional hardware which the already numerous deployed IEEE 802.11-devices help to fulfill. Also vast amounts of solutions [37, 59, 91, 113] based on IEEE 802.11 suggest this to be a useful option for this type of service.

RSS as Distance Indicator As the main indicator for distance only the received signal strength as provided by IEEE 802.11 is used. That means noise and the derived signal-to-noise ratio (SNR) is neglected. The reason is that the noise attribute is not provided by every AP thus it could only be obtained from specific devices. Also Bahl et al. [4] discovered that RSS is a stronger function of location than SNR because the latter is impacted by random fluctuations in the noise process.

Trilateration for Positioning As discussed in section 2.1.1 “Trilateration” RSS is not seen as a very accurate basis for trilateration. The reason it is used here, is that it only acts as a convince method for initial node alignment, not as a live positioning technique. The results of trilateration are subject to manual review and adjustment and usually are not used raw. The details of this method can be further examined below in section 3.1.2 “Alignment of Nodes”.

Propagation Models for Distance Measurement As discussed in subsection 2.2.3 “Models for Indoor Attenuations” propagation models only estimate path loss on a very general level. That is why site-specific parameters are introduced, though they also may only apply to those circumstances researchers derived them from (specific office buildings for example). However, with the presented system a reference level can be measured that can directly be translated into a form used to adopting the model to the exact conditions the systems is used in. A highly customized model is to be expected with only a small effort and without the administrator needing to know any domain-specific details.

3.1.2 Components

In the following subsections the design of the main components are discussed which are ANs, SCS and client.

Active Node

An active node (AN) is a networking device with built-in IEEE 802.11 compatible wireless network radio unit and capable of running an embedded Linux distribution. Due to this requirements it is possible to use traditional Internet routers as well as other existing devices which can be upgraded for the purpose, herewith limiting the need for additional hardware. To qualify, the available device must be able to run a simple web server and at least one dynamic scripting language or framework. The WLAN feature has to be accessible either directly or preferably via a command line front-end so that hardware informations can be accessed and WLAN surveys can be initialized. A survey returns all other WLAN APs in range with service set identifier (SSID), basic service set identifier (BSSID) and RSS in decibel-milliwatts (dBm) among others (see tab. 2.8, p.42). The component diagram of an active node can be seen in figure 3.3.

Most consumer grade APs or WLAN-Routers are type of devices which qualify under these requirements. This is fortunate because they are widely available, cheap, small low-powered electronics and ready-to-use but still with enough computational power to provide a good base for the services to run on the device.

Web Services ANs have to be accessible through Hypertext Transfer Protocol (HTTP) at least over the local area network (LAN) although port-forwarded access over the wide area network (WAN) is a viable alternative if the circumstances would require so. The used web services for network communications are REST-style and XML-based. The reason to use Representational State Transfer (REST) is that it is a widely supported, light-weight architecture style that bases its paradigms on the thoroughly specified HTTP. It is therefore very easy to implement even with simple scripting languages like shell script and rudimentary web servers without considerable overhead. The decision to use Extensible Markup Language (XML) is basically based on the returned output of command line tools that may be very long and contain arbitrary characters which may collide with syntax of certain structures, XML's `CDATA` which allows arbitrary content and therefore may collide, is required. Also it is easier to read in opening and closing tags than with e.g. the property-value syntax of JavaScript Object Notation (JSON). Because support of both JSON and XML are wide spread and the former presented no additional benefit, the choice was to use the more readable and stable format, i.e. XML.

API Before an active node can be used it must be registered with the SCS. To enable this, two services are provided: `ping` and `iwinfo`. The former simply indicates if the web server and the device itself are working and online, the latter provides with its parameterless call hardware and software related informations so that the server knows, i.a. how many radio chips and virtual local area networks (VLANs) are present. This is important, to be able to recognize the different WLAN frequencies that the device supports and to aggregate all SSIDs emitted from a single radio chip. During the measurement phase `iwinfo` can be called with an adapter name (the WLAN driver's abstraction for a radio

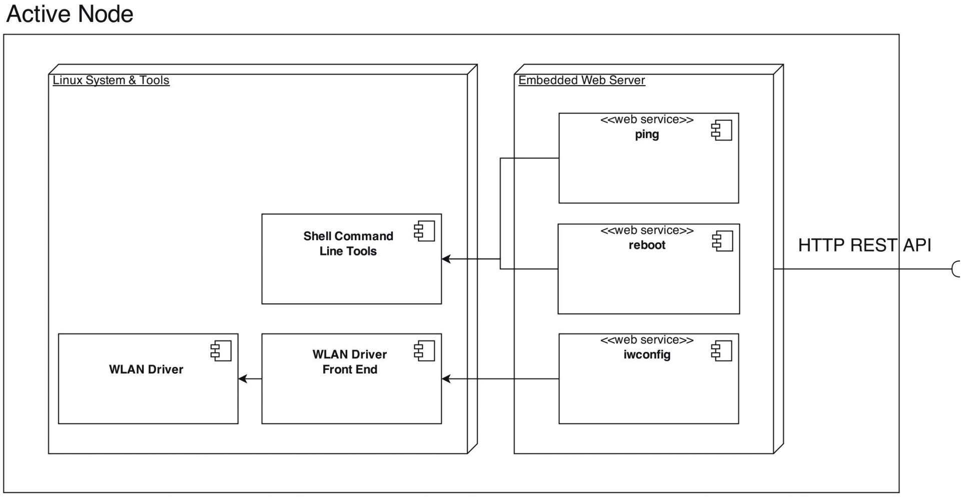


Figure 3.3: Active Node Component Diagram

chip or VLAN) to kick off a survey. As maintenance feature `reboot` can remotely reboot the device in case of software malfunction. A summary of the API can be reviewed in table 3.1 and the full description can be found in appendix B.1.

Service Name	Method	Description
`ping`	GET	Returns a simple response to check if the server is online
`iwinfo`	GET	Returns hardware information and can start a survey
`reboot`	GET	Reboots the device

Table 3.1: A summary of the published web services

Security To restrict the access of ANs to only authorized clients, a couple of techniques can be applied. Although embedded networking devices have limited resources some simple but effective methods are usually available. First of all, most web servers support basic access authentication which is a simple user name/password authentication system specified with the HTTP. It requires the request to have these credentials base64 encoded in the header. This method by itself is not very secure but combined with a secure HTTPS channel with TLS/SSL it presents itself as a reasonable security concept. Most even simpler web servers for embedded systems at least provide modules to support HTTPS. To further enhance on this certificate-pinning can be implemented on the server side so that only communication with valid active nodes may be possible to not reveal the credentials accidentally. Certificate-pinning is a security concept where all valid certificates for TLS/SSL handshakes are stored on a white list and only communication with parties using those certificates are possible. This will not stand up against opening up to the WAN but if the ANs are kept local the presented methods are adequate means to protect from unauthorized access and secures communications.

Service Control Server

The main part of the system is represented by the service control server (SCS). It manages ANs, the map generation and hosts the services for live positioning and lastly presents a user interface for setup and review. The basic components as observable in-detail in figure 3.4 are: scheduler, evaluation & adaption, node-alignment, map generation, location service.

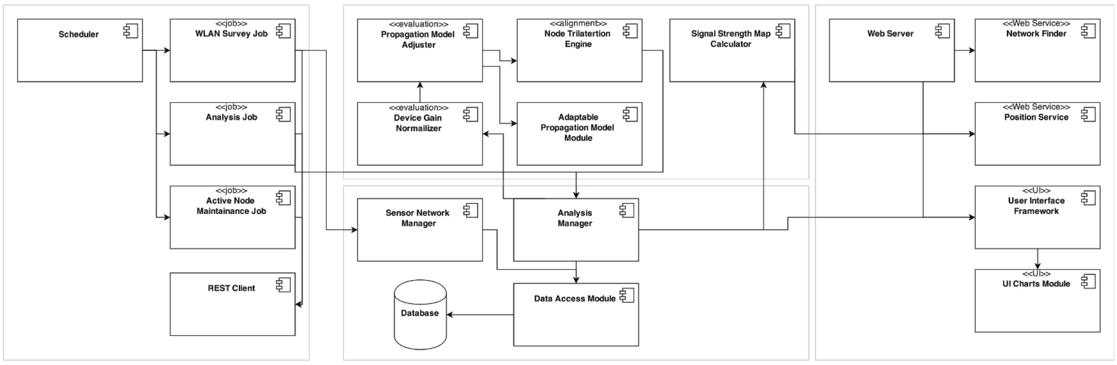


Figure 3.4: Component Diagram of the service control server (SCS)

In figure 3.5 an outline of the location service sequence is shown. The scheduler manages measurement and analysis. An analysis is derived from a certain amount of measurements over a prolonged period of time and generates aggregated statistical data. This is the basis for the site- and device-evaluation so that finally the map with the rough positions of the nodes can be computed and modified by the user. From that the *signal strength map* is generated and used for the location service later on.

In the following all major components of the SCS are discussed.

Job Scheduler Because the system is contentiously adapting itself certain tasks have to be run in continuous intervals (see figure 3.6). First of these is the `survey-job` which is responsible for collecting WLAN surveys of all ANs and storing them in a database. This

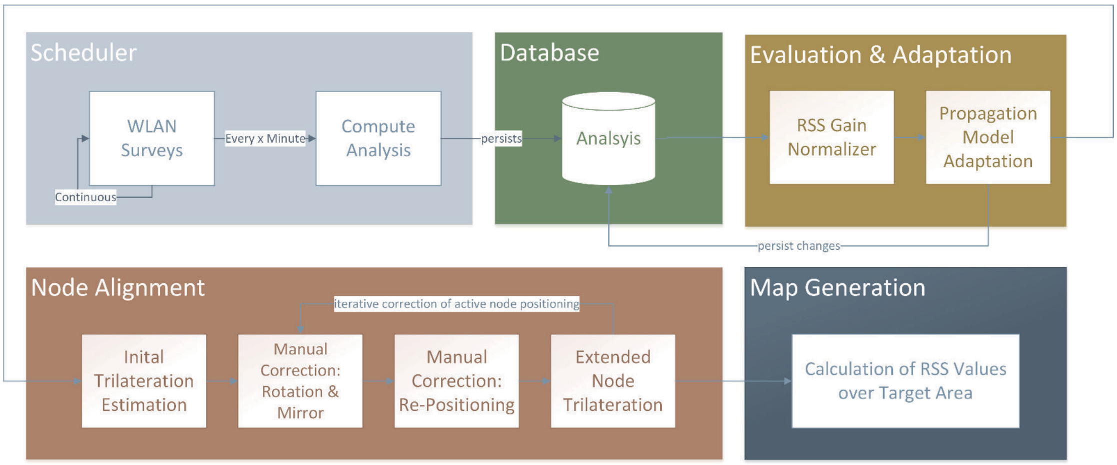

Scheduler
WLAN Surveys
Every x Minute
Compute Analysis
Continuous
persists
Database
Analsyis
Evaluation & Adaptation
RSS Gain Normalizer
Propagation Model Adaptation
persist changes
Node Alignment
iterative correction of active node positioning
Inital Trilateration Estimation
Manual Correction: Rotation & Mirror
Manual Correction: Re-Positioning
Extended Node Trilateration
Map Generation
Calculation of RSS Values over Target Area

Figure 3.5: Diagram of the steps needed to generate the signal strength map

has to be done concurrently for all nodes since the surveys should start at the exact same time due to RSS fluctuations that may distort the data. Then the `analysis-job` is used to aggregate the survey data of a certain time frame and compute base data structures to help with the later site-evaluation and map generation. This includes: calculation of standard statistical properties like mean and median, trends in comparison to older analyses, distinct lists of all extended nodes and basis for the signal map. Additionally as a maintenance feature the `ping-job` records and indicates the online status of all ANs, hence allowing to prevent a loss of time with measurements with some of the nodes not responding and therefore falsifying the readings and analysis.

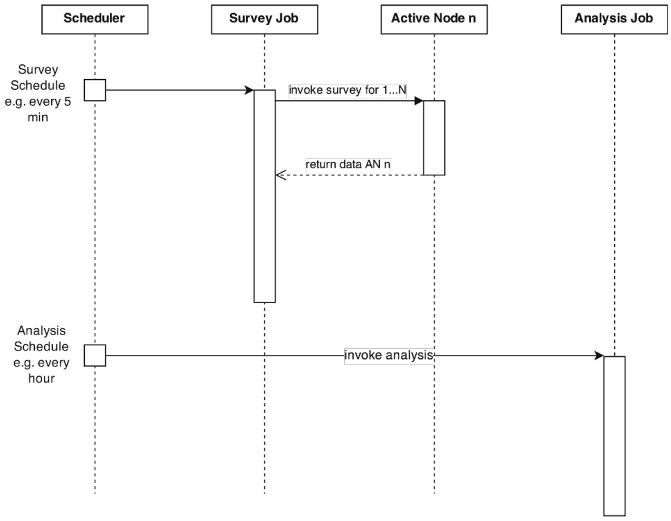


Figure 3.6: Job Schedule Sequence Diagram

RSS Gain Normalizer Due to different hardware and antenna configurations of the nodes and especially the automatic gain control implemented in these devices RSS values may not coincide for 2 nodes that measure each others RSS. To counter this phenomenon an algorithm can be used that identifies the best factor applied for each AN so that RSS values of opposite nodes have the smallest Manhattan distance i.e. a *nearest neighbor search*. This can be achieved with a RSS distance matrix of all ANs and applying said factor to columns i.e. all readings of a node so that they correspond with the reversed measurement and therefore that

$$\forall i, j \in AN \quad f_i \times RSS_{i \to j} \approx f_j \times RSS_{j \to i} \tag{3.1}$$

holds where f_i and f_j are the needed factors applied to node i and j respectively, AN are all active nodes and $RSS_{i \to j}$ is the RSS value measured by node i gathered from node j. This factor can be expressed as the *general device specific gain* and summarizes all radio propagation factors in an practical and easy way. This is important for later distance estimations to minimize error specially for the positioning of the extended nodes with triangulation. Because the only inputs needed are the signal readings this algorithm can work entirely automatic and does not need to be manually set up.

Site Evaluation SEASNIPS uses propagation models to estimate distances from RSS values. The problem with these models is that they are usually very specific to a certain environment (see section 2.2.3 “Models for Indoor Attenuations”). To bypass this shortcoming parameters are introduced by researchers that are derived from empirical observations. These parameters are then categorized to suit different types of sites e.g. offices or open halls. Though a viable solution even this fine grained classification is often too general to be anywhere near precise. This is the reason SEASNIPS includes a site-evaluation module to modify the propagation model to match the exact environment in signal space and therefore enhance the overall distance measurement accuracy.

This is a 2 step semi-automatic process. As a basis the normalized RSS distance matrix is used. Now the positions of each AN is known and thus the actual physical line-of-sight (LOS) distances. These distances are now used as input for the evaluation, the more of theses are provided, the more precise the algorithm can work, which in turn means not all distances are needed. With this data the module now has certain factors and offsets that can be manipulated to alter the propagation model. In the second phase these are now tried out until

$$\forall i, j \in AN \quad d_{i,j:\text{model}} \approx d_{i,j:\text{input}} \tag{3.2}$$

holds where $d_{i,j:\text{model}}$ is the distance calculated from the propagation model with the given RSS value and $d_{i,j:\text{input}}$ is the actual distance declared by the user. This optimization can again formally be described as *nearest neighbor search* with Manhattan distance as metric where the distance of the input value versus the calculated model output must be minimal in all combinations. After the process is done the result is an adapted propagation model that fits the measurements better and thus keeps increasing precision.

Alignment of Nodes With the calibrated system and a calculated distance matrix (see tab. 3.2), SEASNIPS uses trilateration to try to find an initial position of the active nodes. The basis of these calculations is a grid of squares called the *signal strength map*. Each square is a possible position of a node or user. This grid represents the physical outline of the target area and the actual floor plan can be laid over it for easier orientation. The squares are defined with a certain length in meters and thus defining the maximum achievable accuracy. This is also the reason to use such a format — it decreases computational effort while not limiting the overall accuracy of the system when setup with reasonable values.

	AN_1	AN_2	AN_3	AN_4
AN_1	-	5 m	3 m	2.5 m
AN_2	5 m	-	3.2 m	2.3 m
AN_3	3 m	3.2 m	-	0.4 m
AN_4	2.5 m	2.3 m	0.4 m	-

Table 3.2: Example of a distance matrix with four nodes: AN_1, AN_2, AN_3, AN_4. The best case, here depicted, is an Euclidean distance matrix i.e. $\overrightarrow{AN_iAN_j} = \overrightarrow{AN_jAN_i}$

Due to the limitations of a grid the algorithm for trilateration differs slightly from usual approaches. Obviously circles cannot be accurately drawn onto a grid but never the less, can be approximated. This is done with a midpoint circle algorithm which efficiently draws circles onto rasters (see figure 3.7, left and center). Now only in good cases the circles of trilateration actually meet — usually they only form sectors — so a simple

Gaussian distribution around each point in the grid is calculated that spreads a certain distance in all directions. This is pictured in figure 3.7, right diagram, where the fading opacity represents the lower probability of position. So for example the points represented with a dark blue color have a probability that the AN is positioned here of 0.8 and with middle and light blue 0.3 and 0.1 respectively. This is of course depended upon the used distribution and spread discussed in the implementation section but generally it can be said the further the spread the more positions can be found but also location confidence decreases. To find the most likely position a search algorithm finds the point with the highest combined probability of all participating ANs.

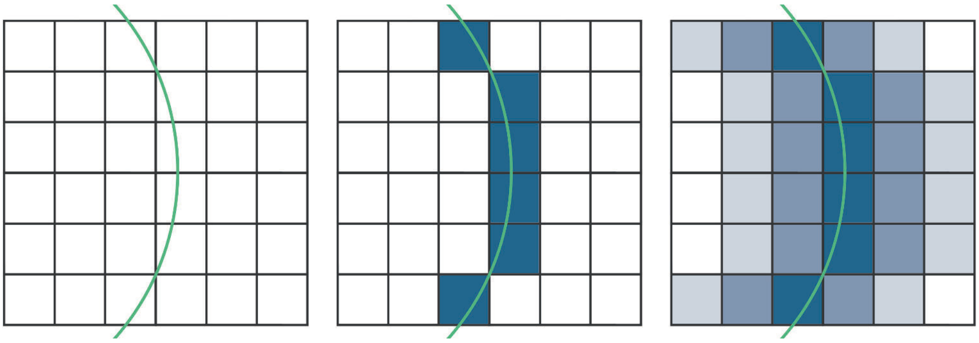


Figure 3.7: Schematic on how a circle is drawn for trilateration in Seasnips

Unfortunately it is not possible to guess the exact positions without additional information, but relative locations with unknown orientation of the $x-$ and $y-$ axis and rotation. These can later be modified to fit the actual physical conditions. To accomplish this the following procedure is used (cf. figure 3.8):

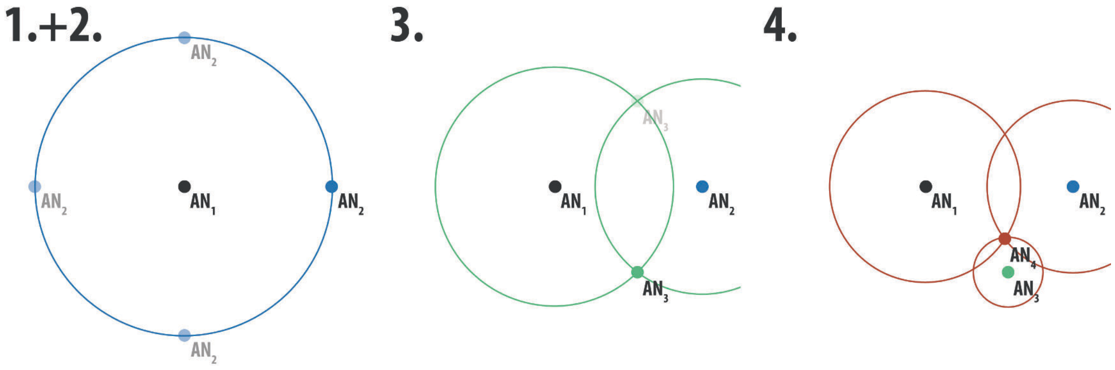


Figure 3.8: Seasnips process of guessing positions of nodes

1. A randomly selected active node (AN), AN_1 is positioned at the center of the base grid i.e. the *signal strength map*.

2. Another AN, AN_2 is randomly selected and from the distance matrix the distance is gathered for $\overrightarrow{AN_1AN_2}$ and a circle is drawn from AN_1 with the radius being the distance. AN_2 could be at any of the unlimited positions of the circle, but to simplify a random location at one of the cardinal directions of the circle is chosen. This is the first guess that may have to be corrected by rotation around AN_1.

3. A third random AN, AN_3 is picked that is seen by AN_1 and AN_2 and a circle is drawn from both of these points with the respective distances to AN_3. Now 2 possible positions for AN_3 are available. This also has to be randomly decided and adjusted by the user later on with mirroring of the resulting graph along one of the axes.

4. After that, three initial positions are guessed and from there on every additional AN can be computed with trilateration.

Now an initial possibly skewed graph of likely positions is available. If this is the case two tools are available to the user to adjust the graph to its actual approximate form: rotation and mirroring. Rotation fixes the first guess along the circle in step 2 and mirroring the second guess in step 3 where one of 2 points are chosen. An example of such correction can be seen in figure 3.9. Aiding this process is a an overlay of the floor plan of the target area. Due to imprecisions in the whole procedure and the discrepancy of the signal space versus the actual physical space some nodes may not be positioned entirely correct, so it may be required to slightly shift some of the ANs. This is done with a simple graphical editor.

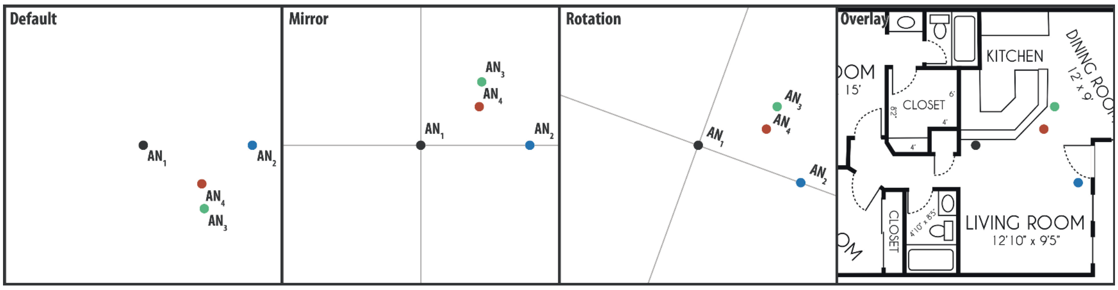
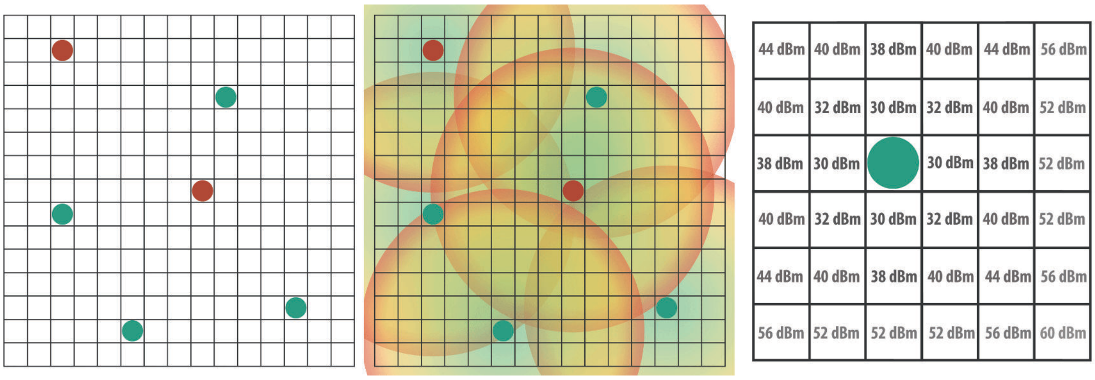


Figure 3.9: Example of manual adjustment of the guessed graph: first it will be mirrored along the x-axis, then rotated with 20° and finally a floor map will be overlaid to give it physical context

To later aid in better and finer grained localization, the positions of all extended nodes (ENs) are determined with the same trilateration procedure as described above. As already mentioned ENs are regular WLAN APs that happen to be in or around the target area. If the result of this trilateration is not satisfying the whole process can be repeated with modified positions of the active nodes. As further optimizations to omit volatile ENs that likely change positions unexpectedly, like for instance mobile 3G routers, a blacklist manages all nodes that are ignored during this whole process.

Signal Strength Map Generation When the positions of all ANs and ENs are calculated the map can now be populated with signal strength values. These values are derived with the adapted propagation model discussed earlier. The wave propagation is simply modeled as with an isotropic antenna (see section 2.3.1) which translates to a circle around each node in a 2D planar (see figure 3.10). The maximal distance is bounded by a configuration value. Usually all nodes are modeled with the same unchanged emitting power but if discrepancies are noticed single nodes can be adjusted to slightly change their modeled power gain if e.g. specific long range antennas are used.



44 dBm	40 dBm	38 dBm	40 dBm	44 dBm	56 dBm
40 dBm	32 dBm	30 dBm	32 dBm	40 dBm	52 dBm
38 dBm	30 dBm	●	30 dBm	38 dBm	52 dBm
40 dBm	32 dBm	30 dBm	32 dBm	40 dBm	52 dBm
44 dBm	40 dBm	38 dBm	40 dBm	44 dBm	56 dBm
56 dBm	52 dBm	52 dBm	52 dBm	56 dBm	60 dBm

Figure 3.10: Example of ANs and ENs in a grid. In the middle figure the estimated radio propagation circles are drawn and in the last a detail of the calculated RSS values with the adapted propagation model is shown (to save space the minus is omitted before every dBm value).

Location Service Finally when the *signal strength map* is annotated with RSS values the base for the positioning service is ready. Two services now can be used by clients to estimate a position (see appendix B.2). Because the system can handle multiple mapped areas simultaneously (e.g. various floors of a building) the first service returns the area that best matches the given SSID and BSSID list from the clients measurement. Next the selected area with all the RSS readings are passed to the second service which then returns the most probable position.

To identify the most likely position Nearest Neighbor (NN) is used (see section 2.1.2 “Nearest Neighbor”). The reason no specified version of this algorithm is needed is that usually the proportions of a square in the grid of the *signal strength map* is so small (reasonable values are 10 × 10 or 20 × 20 cm) that it makes any benefits of k nearest neighbour (KNN) futile. In conclusion the position service works like the online phase of *fingerprinting* (see section 2.1.2 “Scene Analysis/Fingerprinting”)

Positioning Client

No specific requirements are expected from positioning clients. They only need to be able to start a WLAN survey and pass the data to the SCS. Therefore they need a radio module, an antenna and a HTTP-client. A distinction can be made between active clients and passive clients: active clients present their own user interface showing the user where he is on a map n relationship to the device and can start a location request on their own; passive clients are much like ANs such as simple devices with a with a certain REST service installed to them. Passive Clients cannot start a positioning themselves, the SCS is needed for this. A rich user interface is provided with the server to support this positioning with passive clients.

3.2 Implementation

In this section the details and issues of the implementations are discussed. All used technologies and the logic behind these decisions is explained.

3.2.1 Embedded Devices

As already stated regular consumer grade WLAN ANs are used for active nodes and passive clients. To be able to customize these devices a custom operating system or firmware needs to be installed. Fortunately there is already a big community dedicating themselves to creating open software, so plenty of choices are available. Due to high compatibility and extensibility a Linux-kernel based system was selected. So the choice revolved around the specific distribution. Well known projects that fall in this category are: *OpenWRT*, *DD-WRT* and *Tomato*. OpenWRT [90] was chosen since it is fully open source, has a considerably large community active community and is aimed at the developer not the user i.e. provides many tools like its own package manager `opkg` and does not focus on ready-to-use products. It also has a very wide spectrum of compatible devices.

One of the biggest challenges in current embedded network devices is to fit all of the functionality of a Linux distribution in a small memory capacity usually not bigger than 4 or 8 megabyte (MB) in current middle class devices (2014). Due to the system not benefiting greatly from faster HTTP response times, processing power and RAM is not an issue. Fortunately OpenWRT has most needed components already installed in its most basic configuration. All requirements stated in section 3.1.2 “Active Node” are fulfilled out-of-the-box:

web-server OpenWRT has a built-in web-sever, `uhttp` which is internally used for the configuration user interface, called `luci`. It was developed from the ground up for this distribution specifically.

dynamic pages With Common Gateway Interface (CGI) dynamic pages can be served supporting i.a. *shell script* and *Lua* run in-process. Also additional modules are available to support more complex frameworks like PHP, but these usually over-tax the memory budget.

Transport Layer Security (TLS) support Over the package manager `opkg` the TLS module `uhttpd-mod-tls` can be installed for the `uhttp` web server. It can also be mentioned that it supports hard coded basic authentication.

WLAN driver access To provide easy access to all necessary WLAN functionalities, OpenWRT developers created a command-line front-end namely `iwinfo`. It supports all features needed with a unified output structure. This module is either pre-installed or can be added through the package manager `opkg`.

The process of installing OpenWRT is simple with most devices. Through the online documentation the specific build for each device is provided which just needs to be uploaded through the standard firmware-upgrade feature of the vendor user interface. Also other methods exists if the former is not possible, though none where required during the installation processes performed during this work. Various versions of OpenWRT are used since not all of them are supported for all devices; these are `10.3 Backfire`, `12.9 Attitude Adjustment` and the latest stable version at the time of writing `14.7 Barrier Breaker`.

All available services provided by an AN are implemented in *shell script* and *Lua*. The reason is that personal experience has shown that some devices only work correctly with one or the other. For easier setup a shell script was created that installs all needed packages automatically.

Devices

The used devices can be examined in table 3.3 in form of images in figure 3.11. The main model used as active node (AN) is TP-LINK's TL-WR710N. The reason is that they are small, cheap and have the power supply built in, limiting the amount of cables. Also it prevents discrepancies when all devices share the same hardware. As passive client TL-MR3020 was chosen because it can be powered by an external battery capable of providing a 5V line (basically every USB battery pack) to be autonomous. It is also very small and only marginally more expensive than the TL-WR710N. Additionally a surplus WRT54GL was also tested to see the differences in RSS values for different devices.

Model	Vendor	Version	Flash/RAM	Chipset
TL-WR710N	TP-Link	v1.2 (EU)	8 MB/32 MB	Atheros AR9331
TL-MR3020	TP-Link	v1.9	4 MB/32 MB	Atheros AR9331
WRT54GL	Linksys	v1.1	4 MB/16 MB	Broadcom BCM5352

Table 3.3: An overview of all used devices [90]

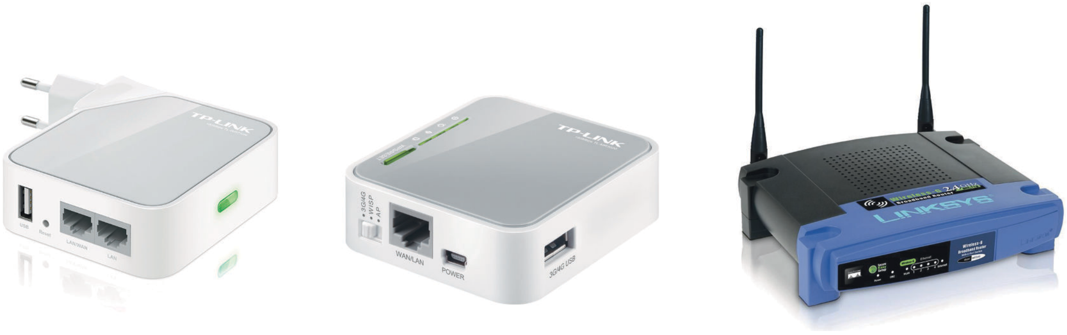


Figure 3.11: From left to right: TL-WR710N, TL-MR3020, WRT54GL [53, 108]

3.2.2 Server

The following contains descriptions of important details of the implementation of the server. It has a similar structure as subsection 3.1.2 “Components”.

Technology Stack

The server is written in Java and JavaScript. It is implemented as Java Servlet version 2.5 to maximize compatibility as a wide range of Servlet containers exist (e.g. *Apache Tomcat 7*, *Glassfish 2* or *Jetty*) i.e. a server that can run Servlet Java programs. Servlet is a very lightweight and flexible technology producing less overhead with full control on the low level side of HTTP. As REST specification Java API for RESTful Web Services (JAX-RS) 2.0 is used with the Jersey implementation in version 2. In combination Jackson 2 is responsible for JSON handling which is also used for persistence. As database Apache CouchDB was chosen, a NoSQL-type document store which is accessed through a HTTP API and uses MapReduce with JavaScript as query language. For convenience, Ektorp is applied as a ready-to-use persistence mapper for CouchDB. The scheduler framework Quartz takes care of all task run in background. Finally the open source library Colt, which stands for high performance scientific and technical computing developed at Conseil Européen pour la Recherche Nucléaire (CERN) is employed to support with some technical calculations.

As front-end technology the web application framework AngularJS is used which enables the user interface (UI) to completely be generated in the client’s browser with only simple REST APIs for communications. Additional UI libraries extended the functionality of AngularJS with e.g. the ability to create charts or render a 2D canvas. To keep the visual design consistent the Cascading Style Sheets (CSS) framework Bootstrap in version 3 is used.

Maven is responsible for build and dependency management. It also helps with deployment and automatic testing with JUnit. A summary of the used versions and licenses can be examined in table 3.4.

All used technologies are either open source or free to use for private and educational purposes, well tested and ready for production.

User Interface

The UI is a single-page JavaScript web-page powered by AngularJS. This framework supports the model-view-controller (MCV) pattern, features double-data binding, decou-

Name	Version	License	Note
Java	7	GPL	
Apache CouchDB	1.5	Apache License 2	document store
Apache Tomcat	7	Apache License 2	Servlet container
Apache Maven	3	Apache License 2	build management
Java Servlet	2.5	CDDL	
Jersey	2.6	CDDL & GPL 2	JAX-RS 2.0
Jackson	2.3	Apache License 2	JSON parser
Ektorp	1.4	Apache License 2	persistence mapper
Quartz	2.2	Apache License 2	scheduler
Colt	1.2	CERN & LGPL	scientific computing
AngularJS	1.3	MIT License	highcharts-ng, ui-bootstrap
Highcharts	4	Non-commercial	JavaScript charts library
PaperJS	0.9.20	MIT License	JavaScript canvas library
Bootstrap	3.3	MIT License	Flatly Theme

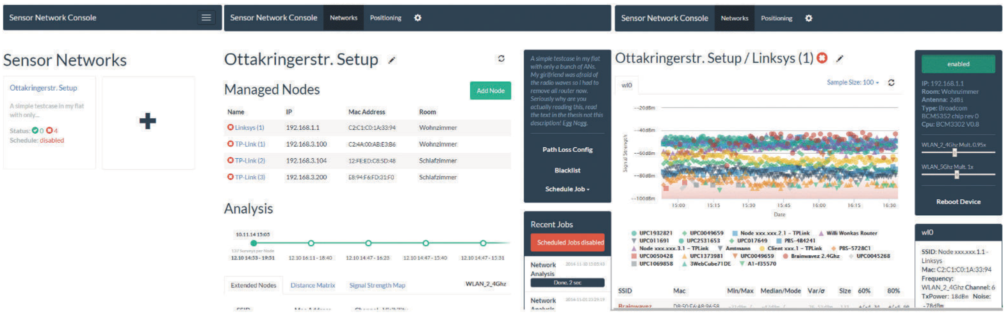
Table 3.4: Technology stack of the SCS

ples the client side from the server side with only using REST APIs, improves testability with decoupling Document Object Model (DOM) manipulation from application logic and provides dependency injection for components among others.

The main screen is a list of *Sensor Networks* i.e. mapped areas (so floor 1 to 4 from a building for example). Each *Sensor Networks* has a list of ANs, a series of analyses, a job status list and navigation to other important features like propagation model adaption. For each AN a detail page can be viewed where the current measurements are depicted. All this can be seen in figure 3.12

Measurement Strategy

As shown by Parameswaran et al. [87], received signal strength is a very fluctuating indicator on the network interface controller (NIC) level. That means that a burst of multiple surveys can result in outliers that are beyond 3 dBm in some cases (more than half or double signal strength) or sometimes in not receiving signals from an AP at all even if it is active. To counter this problem the sample size is drastically increased. Every survey initiated by the SCS is a series of at least five single surveys with a five



A simple testcase in my flat with only...

Name	IP	Mac Address	Room
Linksys (1)	192.168.1.1	C2:C1:C0:1A:33:94	Wohnzimmer
TP-Link (1)	192.168.3.100	C2:4A:00:AB:E3:B6	Wohnzimmer
TP-Link (2)	192.168.3.104	12:FE:ED:C8:5D:48	Schlafzimmer
TP-Link (3)	192.168.3.200	E8:94:F6:FD:31:F0	Schlafzimmer

10.11.14 15:05
137 Serveys per Node
12.10 14:53 - 19:51
12.10 16:11 - 18:40
12.10 14:47 - 16:23
12.10 14:47 - 15:40
12.10 14:47 - 15:31

A simple testcase in my flat with only a bunch of ANs. My girlfriend was afraid of the radio waves so I had to remove all router now. Seriously why are you actually reading this, read the text in the thesis not this description! Egg Nogg.

IP: 192.168.1.1
Room: Wohnzimmer
Antenna: 2dBi
Type: Broadcom BCM5352 chip rev 0
Cpu: BCM3302 V0.8

SSID: Node xxx.xxx.1.1 - Linksys
Mac: C2:C1:C0:1A:33:94
Frequency: WLAN_2_4Ghz Channel: 6
TxPower: 18dBm Noise:

SSID	Mac	Min/Max	Median/Mode	Var/σ	Size	60%	80%

Figure 3.12: Different views in the server's UI

second delay time between each of them. This combination of configurations could be identified as best trade-off of duration (effort) versus confidence in measurements. This is a good measure to balance outliers and missing APs in the resulting statistics. There is no algorithm that filters out outliers because these also are part of the distribution and omitting them could skew the outcome. Because Quartz supports CRON expressions which are highly flexible, intervals can be adjusted arbitrarily with even sophisticated rules per mapped target area. Experience has shown that data quality benfits from very short intervals (every couple of minutes). This enables to generate a extensive sample size in short time of easily around 500-1000 measurements after only a few hours. The result can be seen in figure 3.13 where a graph with a sample size of over 1500 is depicted.

As main indicator, the statistical mean was selected to best represent the summary of the sample size. In systems where only a small amount of samples are measured the median usually is the best choice, but here the mean is slightly more accurate (mean and median are usually within 1 dBm) since medians are only integer values — the highest precision provided by most devices.

RSS Gain Normalizing

As described in section 3.1.2 "RSS Gain Normalizer" a phase in the positioning system is to calibrate each node's RSS values with a factor so that $RSS_{i \to j} \approx RSS_{j \to i}$. To achieve that goal a simple linear nearest neighbor search was developed. For correct results the factor of a random picked node has to be fixed to 1.0 while all the others loop from a set minimum to a maximum with a given step-width. These steps revolve around the starting point i.e. 1 because the best value is expected to be near 1 not near the extremes. Practical values have been shown to be 0.9 – 1.1 as factor. The initial step of 0.001 adjusts accordingly if it exceeds a threshold depending on the maximum estimated

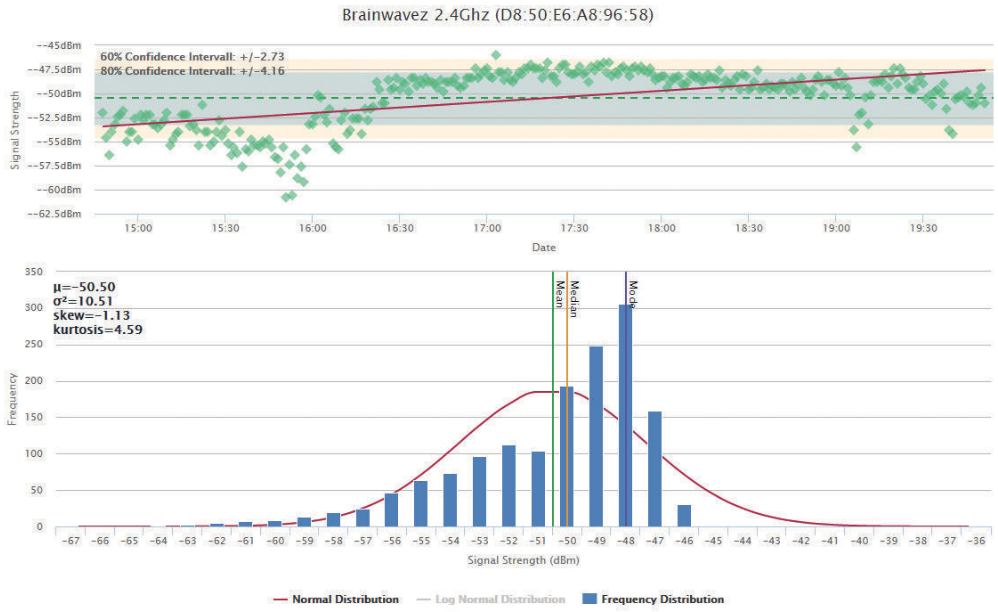


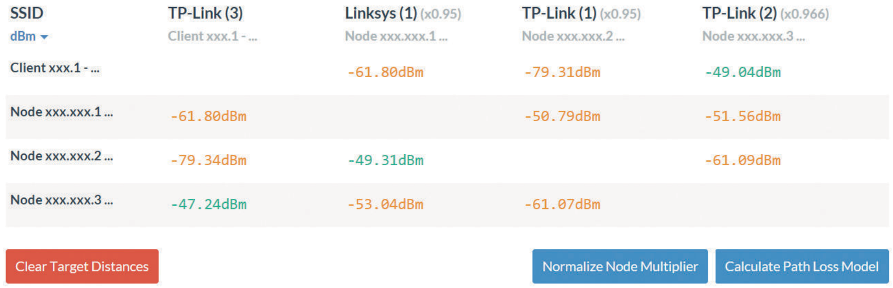
Figure 3.13: This is extracted from the actual administrative UI: the top scatter charts shows RSS of a single AP over a time axis; the red line is the regression and the plot-bands display the 80th and 60th confidence interval; the bottom bar-graph is the frequency distribution with a Gaussian cumulative distribution function (CDF) plotted over it

iterations. Iterations raise exponentially if more ANs are present. In each loop the sum of all Manhattan distances are computed and compared to the “best-so-far” value. This is further performance optimized by multi-threading for each fixed AN because these computations are independent of each other. Further optimization are not required since this algorithm is able to successfully optimize four ANs on a relatively moderate *Intel Core i5-2520M* in under four seconds. As already stated, more ANs would limit the iterations so no exponential raise is to be expected since a threshold caps computational time. In figure 3.14 a matrix resulting from a normalization can be examined.

Evaluating the Site

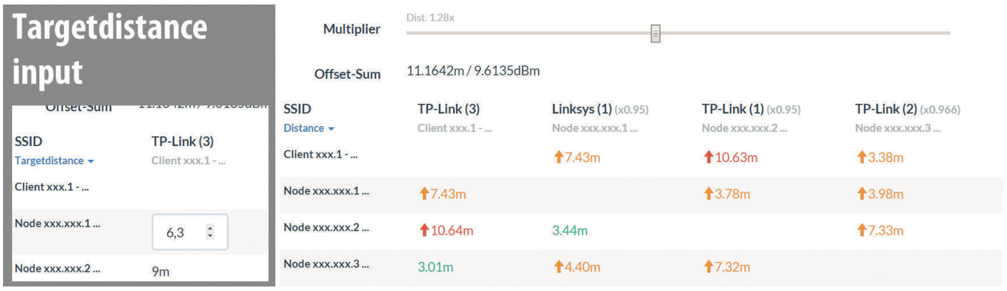
Site adaption is as already described in section 3.1.2 “Site Evaluation” a 2 phase procedure. First the physical distances between all ANs need to be estimated and then inserted into a target distance matrix as seen in figure 3.15. It then immediately shows with colors and trend indicators how well the model is currently adapted for the given scenario.

The distances are calculated with the *ITU indoor propagation model* (see section 2.2.3) since it generally showed good results with higher signal strengths though it had problems



SSID dBm ▾	TP-Link (3) Client xxx.1 - ...	Linksys (1) (x0.95) Node xxx.xxx.1 ...	TP-Link (1) (x0.95) Node xxx.xxx.2 ...	TP-Link (2) (x0.966) Node xxx.xxx.3 ...
Client xxx.1 - ...		-61.80dBm	-79.31dBm	-49.04dBm
Node xxx.xxx.1 ...	-61.80dBm		-50.79dBm	-51.56dBm
Node xxx.xxx.2 ...	-79.34dBm	-49.31dBm		-61.09dBm
Node xxx.xxx.3 ...	-47.24dBm	-53.04dBm	-61.07dBm	

Figure 3.14: The RSS matrix with the factors applied to normalize the values. The accumulated Manhattan distances could be decreased to 4.8 dBm from a 15.97 dBm in this case.



SSID Targetdistance ▾	TP-Link (3) Client xxx.1 - ...
Client xxx.1 - ...	
Node xxx.xxx.1 ...	6,3
Node xxx.xxx.2 ...	9m

Multiplier: Dist. 1.28x

Offset-Sum: 11.1642m / 9.6135dBm

SSID Distance ▾	TP-Link (3) Client xxx.1 - ...	Linksys (1) (x0.95) Node xxx.xxx.1 ...	TP-Link (1) (x0.95) Node xxx.xxx.2 ...	TP-Link (2) (x0.966) Node xxx.xxx.3 ...
Client xxx.1 - ...		↑7.43m	↑10.63m	↑3.38m
Node xxx.xxx.1 ...	↑7.43m		↑3.78m	↑3.98m
Node xxx.xxx.2 ...	↑10.64m	3.44m		↑7.33m
Node xxx.xxx.3 ...	3.01m	↑4.40m	↑7.32m	

Figure 3.15: On the left an example of the UI for inputting the target-distance is shown and on the right the currently calculated propagation model distances and indicators for the magnitude of offset to the input target values.

with low RSS as also seen by the COST-231 [23, p.176] committee. Therefore the following parameters for adapting the model are used:

environmental model The empirical parameters provided by the ITU model itself. See appendix A.2 for a full list.

offset Added in meter to distance calculation and thus is the minimum distance in the model. This is to prevent unrealistically small distances with very high received signal strength.

bound From which dBm path loss value the bound-factor is used. This is the limit where the original model starts to estimate wrong distances.

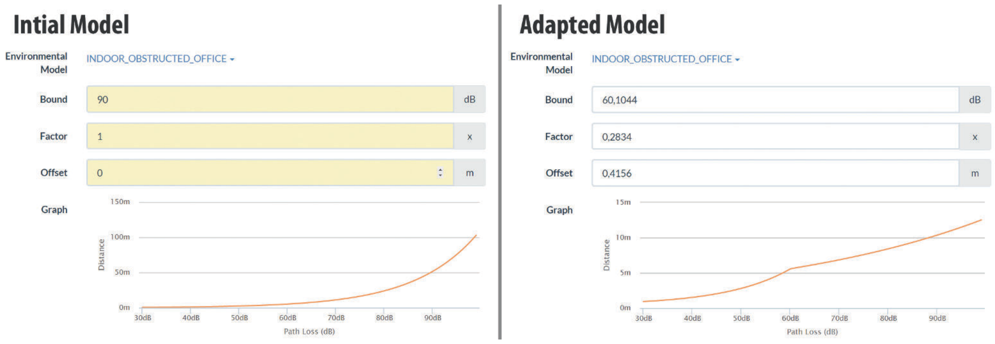
bound-factor The factor which the dBm input values are multiplied by if they are greater than the bound. This factor is needed to adjust the incorrectly estimated distances to match the real values.

general multiplier A general factor with which all distance values are multiplied by.

The application of these parameters in the model can be examined in equation 3.3 where the used adaptive propagation model is described.

$$d_{\text{adapted}} = m_{\text{general}} \times \begin{cases} k_{\text{offset}} + d_{\text{itu}}(j_{\text{bound}} + ((PL - j_{\text{bound}}) \times f)) & b < PL \\ k_{\text{offset}} + d_{\text{itu}}(PL) & b \geq PL \end{cases} \tag{3.3}$$

To automatically find the best configuration a similar algorithm as described in the above section is used, linear nearest neighbor search, but with a slightly different strategy. All the above mentioned variables are looped through within given bounds and step-widths but contrary to the normalization process only a very rough resolution is used in the first iteration (big steps) which are refined with each successive iteration. For example the whole process starts with a step-width of 0.1 and after finding the best match the bounds are narrowed to $d_{\text{min}} \pm 0.1$ and the step width is decreased to 0.01. This continues for a defined amount of iterations until the shortest overall distance down to multiple fractions is achieved. No specific performance optimizations are applied since it runs in under 5 seconds on a *Intel Core i5-2520M* with a precision of 10^{-4}. A resulting plot of the model along path loss values can bee seen in figure 3.16.



Intial Model

Environmental Model: INDOOR_OBSTRUCTED_OFFICE

Bound: 90 dB

Factor: 1 x

Offset: 0 m

Adapted Model

Environmental Model: INDOOR_OBSTRUCTED_OFFICE

Bound: 60,1044 dB

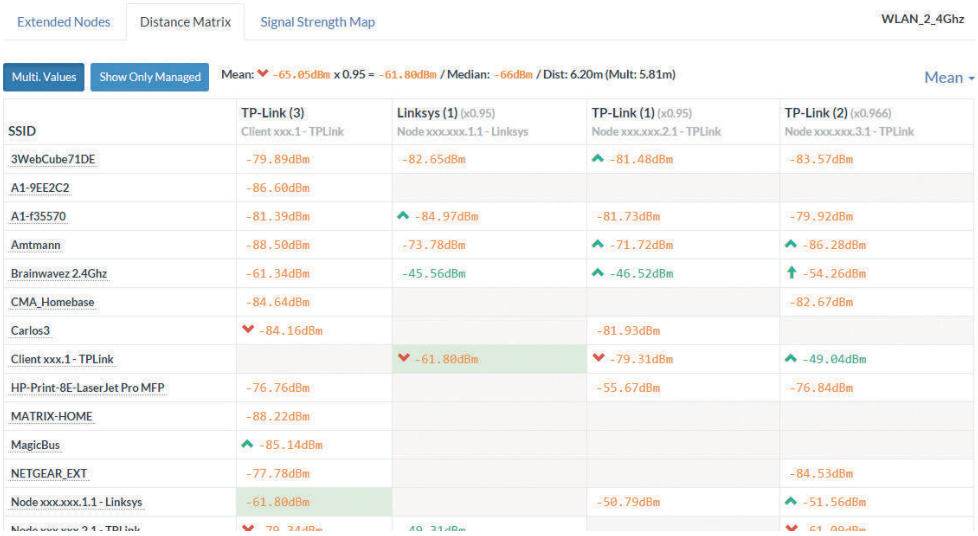
Factor: 0,2834 x

Offset: 0,4156 m

Figure 3.16: On the left the unmodified model is shown with the default function plot; on the right the adapted model after automatic calibration is shown with all used parameters. In this instance the overall distance offset could be decreased to 1.3 m from an initial 33 m

Analysis

An analysis is made after a certain amount of measurements. This is, like the survey strategy, customizable to a high degree with CRON expressions. Analysis present accumulated data from said measurements for the administrator to review (e.g. to see if certain APs changed drastically in RSS) and provide the base for trilateration of the ANs. In figure 3.17 a resulting distance matrix of an analysis can be seen. The rows represent ENs & ANs and the columns are only ANs. Merely EN with a visibility of three or more are visible, because below that number they cannot be positioned using trilateration.



Mean: ▼ -65.05dBm x 0.95 = -61.80dBm / Median: -66dBm / Dist: 6.20m (Mult: 5.81m)

SSID	TP-Link (3) Client xxx.1 - TPLink	Linksys (1) (x0.95) Node xxx.xxx.1.1 - Linksys	TP-Link (1) (x0.95) Node xxx.xxx.2.1 - TPLink	TP-Link (2) (x0.966) Node xxx.xxx.3.1 - TPLink
3WebCube71DE	-79.89dBm	-82.65dBm	▲ -81.48dBm	-83.57dBm
A1-9EE2C2	-86.60dBm			
A1-f35570	-81.39dBm	▲ -84.97dBm	-81.73dBm	-79.92dBm
Amtmann	-88.50dBm	-73.78dBm	▲ -71.72dBm	▲ -86.28dBm
Brainwavez 2.4Ghz	-61.34dBm	-45.56dBm	▲ -46.52dBm	↑ -54.26dBm
CMA_Homebase	-84.64dBm			-82.67dBm
Carlos3	▼ -84.16dBm		-81.93dBm	
Client xxx.1 - TPLink		▼ -61.80dBm	▼ -79.31dBm	▲ -49.04dBm
HP-Print-8E-LaserJet Pro MFP	-76.76dBm		-55.67dBm	-76.84dBm
MATRIX-HOME	-88.22dBm			
MagicBus	▲ -85.14dBm			
NETGEAR_EXT	-77.78dBm			-84.53dBm
Node xxx.xxx.1.1 - Linksys	-61.80dBm		-50.79dBm	▲ -51.56dBm

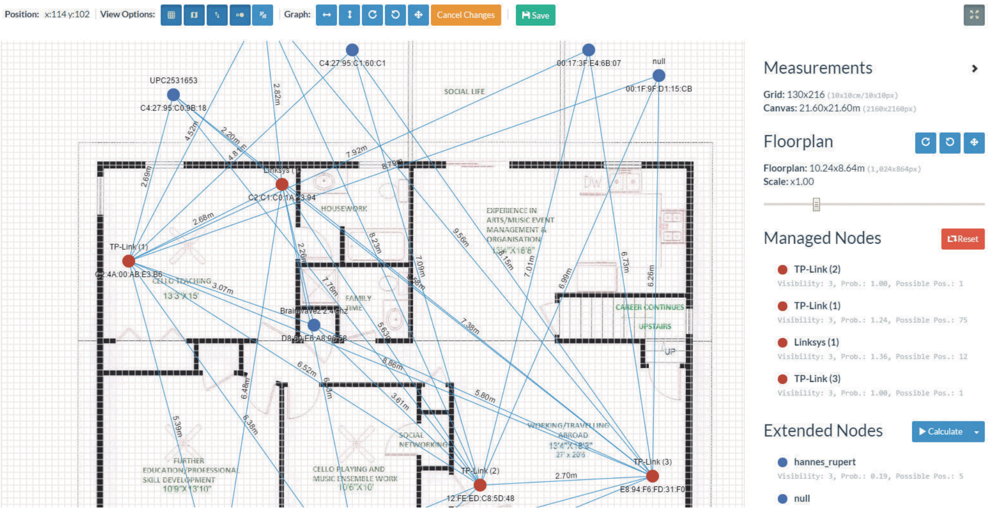
Figure 3.17: The distance matrix after an analysis as presented in the UI

For extended reviews different view option can be selected to show either, mean, median, mode, min, max or the standard deviation. Also the distance calculations with mean and median can be compared. Arrows indicate tendencies compared to older analysis for easier identification of unstable signal sources like mobile ENs.

Visual Node Editor

The next step is the adjusting of automatic trilateration positioning of ANs. To help with this task a floor plan upload feature is available. Now the nodes or the plan can be shifted or scaled to create a decent overlay. After that the ENs can be positioned using trilateration with the click of a button. An example of this is depicted in figure 3.18 where a finished positioning can be seen on a 10 × 10 cm grid. On the right all ANs and ENs in red and blue respectively are listed with further details from the trilateration process to get an idea how confident the system is that a certain position is correct. These are represented with the total visibility, a probability value and the count of all possible positions with the same probability. The extended node calculation lets the user set the option how far the spread used is for the process (as described in section 3.1.2

"Alignment of Nodes") to set what kind of confidence range can be tolerated. On the top, graph manipulation tools can be accessed like rotation, mirroring and panning as well as basic functions like saving.


Position: x:114 y:102

Grid: 130x216 (10x10cm/10x10px)
Canvas: 21.60x21.60m (2160x2160px)

Floorplan: 10.24x8.64m (1,024x864px)
Scale: x1.00

TP-Link (2)
Visibility: 3, Prob.: 1.00, Possible Pos.: 1

TP-Link (1)
Visibility: 3, Prob.: 1.24, Possible Pos.: 75

Linksys (1)
Visibility: 3, Prob.: 1.36, Possible Pos.: 12

TP-Link (3)
Visibility: 3, Prob.: 1.00, Possible Pos.: 1

hannes_rupert
Visibility: 3, Prob.: 0.19, Possible Pos.: 5

null

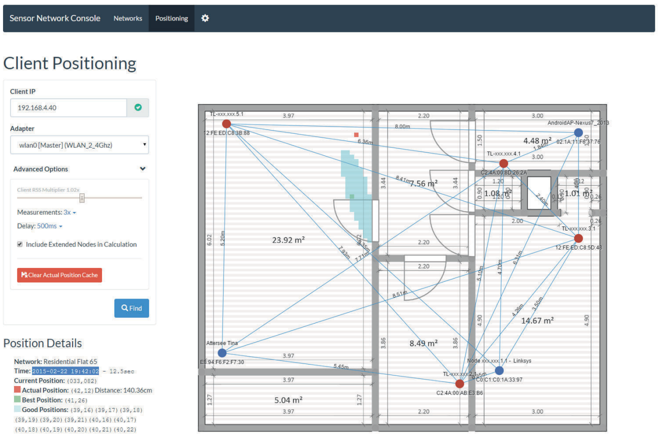
Figure 3.18: The visual node editor where ANs, ENs and the floor plan can be adjusted in multiple ways.

Radio Map Generation and User Positioning

For the radio map, which stores all calculated RSS values for later positioning, a *Sparse Matrix* is used. In a *Sparse Matrix* only set elements consume memory. This helps support large grids i.e. areas, without wasting heap-memory and therefore preventing exceeding it prematurely. Unfortunately sparse-type data-types are not implemented in the Java Development Kit (JDK) thus the implementation of *Colt* (see section 3.2.2) is used. This library offers decent performance with average get/set runtime of $\mathcal{O}(1)$. A more efficient data-type in this case would be a *Trie* due to their faster reading performance, but no production-ready implementation was readily available in Java.

Every element in the matrix stores a map with BSSID as key and the RSS in dBm as value (see figure 3.10). During the location phase 2 services are published: selecting the mapped area and finding the most likely position in that area. The first service just compares all given BSSIDs with the measured ones in the respective mapped area and the best match is chosen. The position algorithm works, as already stated in section 3.1.2 "Location Service", like *fingerprinting* with a normal NN algorithm. For that a list of all set elements is extracted from the *Sparse Matrix* and looped over until the best position(s) are found with Manhattan distance as metric.

In the UI the best and additionally the best 100 positions are displayed as can be seen in figure 3.19. When clicking on the map the actual position can be selected and the system calculates the error offset to it in meters.



Client IP: 192.168.4.40

Adapter: wlan0 [Master] (WLAN_2_4Ghz)

Client RSS Multiplier 1.02x

Measurements: 3x

Delay: 500ms

☑ Include Extended Nodes in Calculation

Network: Residential Flat 65
Time: 2015-02-22 19:42:02 - 12.5sec
Current Position: (033,082)
🟥 Actual Position: (42,12) Distance: 140.36cm
🟩 Best Position: (41,26)
🟦 Good Positions: (39,16) (39,17) (39,18) (39,19) (39,20) (39,21) (40,16) (40,17) (40,18) (40,19) (40,20) (40,21) (40,22)

Figure 3.19: The UI of client positioning.

Position Accuracy Modes An important parameter for the accuracy of positioning is the count of the surveys that are used as a base for the algorithms. Setting it too low may result in very fluctuating and inconsistent results but setting it too high extends the duration of the process too much for it to be practical. In this case the time increases linearly which, as a reference, takes about an additional four seconds every added survey with four ANs in the system. To identify the best setup several measurements were carried out with different settings. As it can be shown in figure 3.20 the best trade-off of time and accuracy is three surveys per positioning attempt; with four it seems that diminishing returns start to begin. Also the general accuracy can be slightly enhanced as seen in the mean offset bar chart.

To classify modes, the three surveys with 500 ms between each will be defined as *high accuracy*, whereas the two and one survey setting will be defined as *medium* and *low accuracy* respectively. As reference, usual durations for these processes are 4.5, 8.4 and 12.5 seconds.

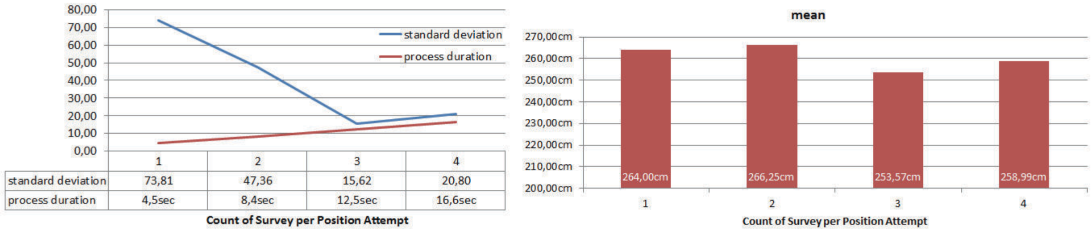


Figure 3.20: On the left the standard deviation of the measurements are shown. The x-axis represents the count of WLAN surveys that are used as a base for a single positing process. On the right the mean offset is shown. The sample sets consists of 15 position processes each.

CHAPTER 4

Evaluation

This chapter presents a comparative performance evaluation of the proposed system *Site Evaluating Active Sensor Network Indoor Positioning System* (SEASNIPS) . It describes the setup in detail and presents and discusses the results. Because the system is based on collecting large sample sizes of received signal strength (RSS) data, the first section tries to interpret its statistical properties and their implications as a base for indoor positioning.

4.1 Observations with RSSI

This section investigates the statistical properties of RSS in live conditions. It is gathered by measurements with the access points (APs) described in section 3.2.1 “Devices”. Interpreting the received signal strength is highly subjective relative to the device and the environment; thus, these observations are only valid for this specific instance with these devices. Some conclusions may be generally valid, although this cannot be guaranteed without more extensive testing. The sample size comprises 2,500 surveys over about 14 hours. The count of the actual readings from certain APs fluctuate because not all of them are consistently found in all the surveys, especially with low signal strengths. This is also very important information to keep in mind, since as shown in figure 4.1, the chance of being discovered decreases drastically if the RSS is below -85 dBm.

An important property is how consistent the RSS readings are with different signal strengths. Intuitively, one would expect lower RSS values to fluctuate more than higher ones. Unfortunately, this could not be observed here, as seen in figure 4.2. According to

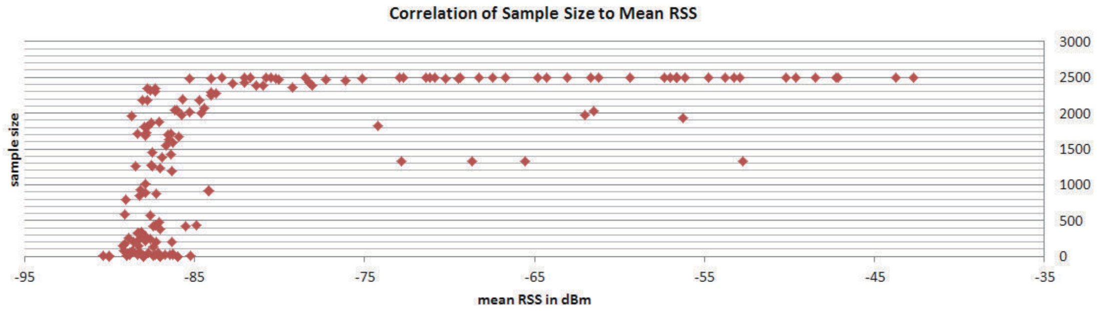


Figure 4.1: Correlation of mean to sample size from the gathered data of Experiment 1.

the trend line, the standard deviation decreases slightly the weaker the signal strength becomes. Furthermore, there are very few outliers after about −75 dBm and some of the highest standard deviation values are above −55 dBm. It also should be noted that the sample size after −85 dBm can be quite low, as mentioned above, and so there is perhaps less of a chance of the standard deviation being high. However, all in all, there is no strong correlation between these two properties.

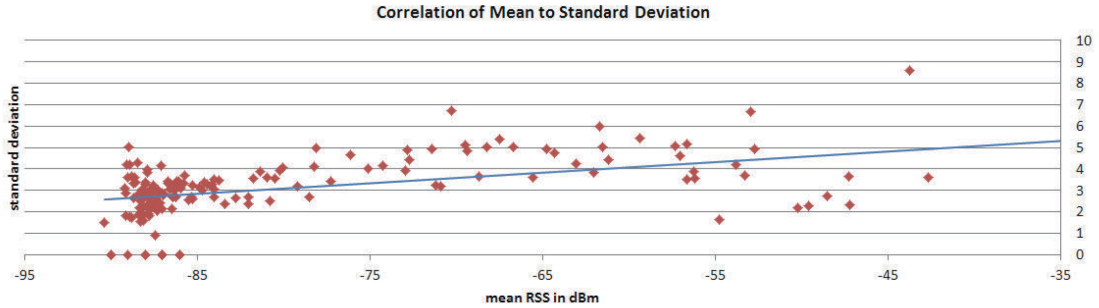


Figure 4.2: Correlation of mean to standard deviation from the gathered data of Experiment 1.

Another indicator of the variation of RSS readings comprises the minimum and maximum values. As can be examined in 4.3, maximum values deviate further from the mean with a stronger RSS, but generally the min-max range is consistent around the mean, but it cannot be used to derive it reliably, as indicated by the many spikes.

A significant design decision for SEASNIPS was that a generalization approach would be used. For this, the mean was chosen and, as can be seen in figure 4.4, the median correlates nearly linearly to it, which gives the mean the advantage of finer facets. The third possibility, the mode, would simplify the distribution too much.

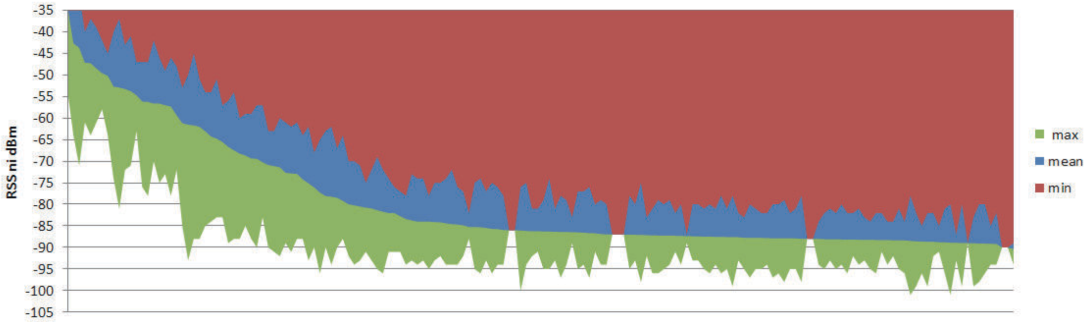


Figure 4.3: Minimum, maximum and mean from the gathered data of Experiment 1.

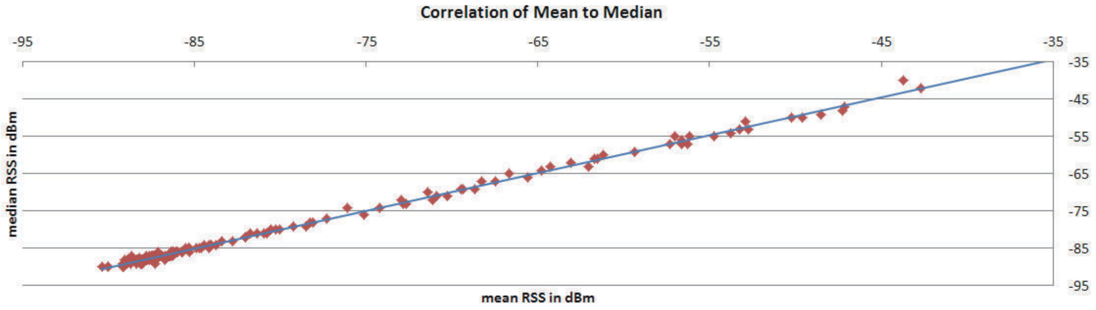


Figure 4.4: Correlation of mean to median from the gathered data of Experiment 1.

This brings us directly to the next important question: what is the underlying distribution? This cannot be answered clearly. It seems that multiple distributions are in play, including a normal distribution. One indicator of whether a sample is normally distributed is that its mean, median and mode are identical, which the graph in figure 4.5 tries to display with the offset (smaller values mean a higher probability of a normal distribution) of the said properties in relation to the mean RSS.

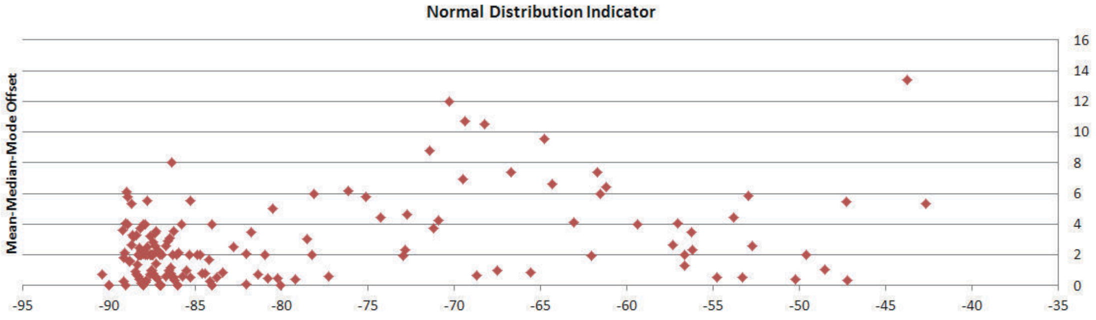


Figure 4.5: Normal distribution indication from the gathered data of Experiment 1.

As can be seen, with the exception of certain outliers, most of the distributions satisfy this requirement. As an example, in figure 4.6 a perfect instance can be examined where the underling distribution is clearly normal, with a slight right skewness.

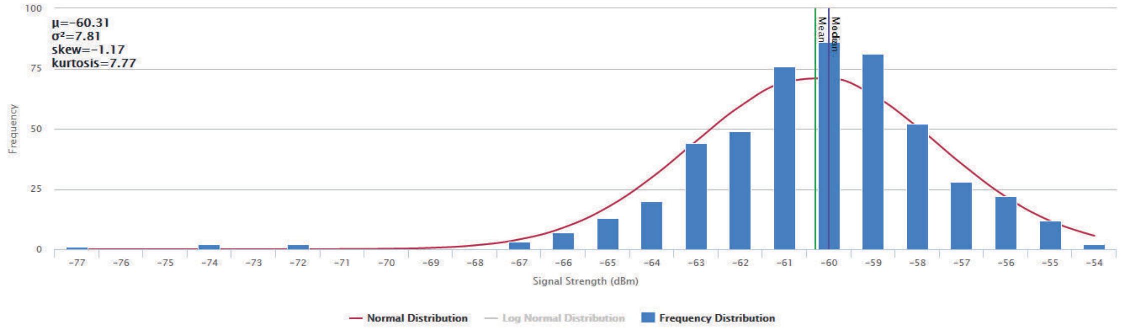


Figure 4.6: Example of frequency distribution showing a normal distribution from the gathered data of Experiment 1.

Another example of one of the outliers from the chart in figure 4.5, the frequency graph can be compared in figure 4.7 where it seems that it is influenced by multiple distributions.

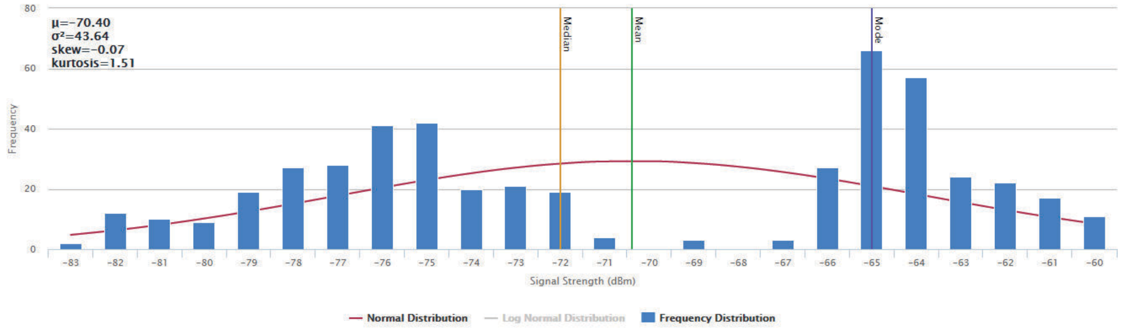


Figure 4.7: Example of frequency distribution showing an unknown distribution from the gathered data of Experiment 1.

It therefore cannot generally be said that RSS is normally distributed, but in some instances it is. However, in these observations a correlation cannot be made between signal strength and type of distribution, as is claimed in the literature [17, 58]

4.2 Methodology

4.2.1 Comparative Product

To compare the results of various metrics to an existing commercial product, the indoor tracking and positioning service indoo.rs was chosen. The company which developed this product was founded in Austria and offers solutions based on RSS fingerprinting with Wireless Local Area Network (WLAN) and proximity with Bluetooth (BT) v4. According to their website, indoo.rs has implemented indoor navigation services in several airports, shopping malls and for various events. They provide a measurement tool as well as a simple showcase app for Android and iOS which can by synchronized via their servers. The system will be used in WLAN fingerprinting mode only, which is seen as the main competing approach to Seasnips . It is expected that indoo.rs will include a highly optimized positioning algorithm making it a valid reference for this experiment.

4.2.2 Method

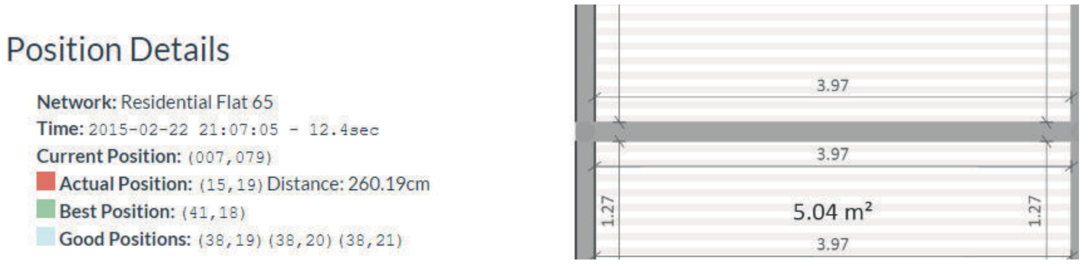
In the following, the methods and the general approach of the experiments performed are described. First, a floor plan of the target area is created. As a tool, the online planer *Home Styler* from Autodesk in its current beta version is used. The fundamental measurements are carried out with the help of a digital laser-measuring tape by *Bosch* to minimize the measuring faults.

During the RSS measurement and positioning phase, the site is kept as consistent as possible, which means that, e.g., doors and windows stay in their respective states. Human and “animalistic” interference is omitted as far as possible.

Seasnips

The active nodes (ANs) are distributed homogeneously over the target area. The system is set up and a measuring phase of at least three hours is carried out. Next, the usual propagation model calibrations are performed and the ANs as well as the extended nodes (ENs) are positioned over the floor plan in the signal map representation in the user interface (UI). In this setup, ENs is only used if its position can be physically verified - trilateration estimations will not be trusted.

In the positioning phase, several defined positions are selected. The client device will be placed at these locations, which are measured with the laser tape to minimize errors resulting from the inconsistent setup. During the actual surveys, no person should be in the vicinity. For accuracy mode, *high* (see section 3.2.2 “Position Accuracy Modes”) is chosen, which takes about 12 seconds per positioning process. Each positioning will be done consecutively at least three times. In the UI, the actual position can be defined and the offset is calculated automatically as can be seen in figure 4.8.



Position Details

Network: Residential Flat 65
Time: 2015-02-22 21:07:05 - 12.4sec
Current Position: (007,079)
🟥 **Actual Position:** (15,19) Distance: 260.19cm
🟩 **Best Position:** (41,18)
🟦 **Good Positions:** (38,19) (38,20) (38,21)

Figure 4.8: Example of a positioning with offset to its actual position.

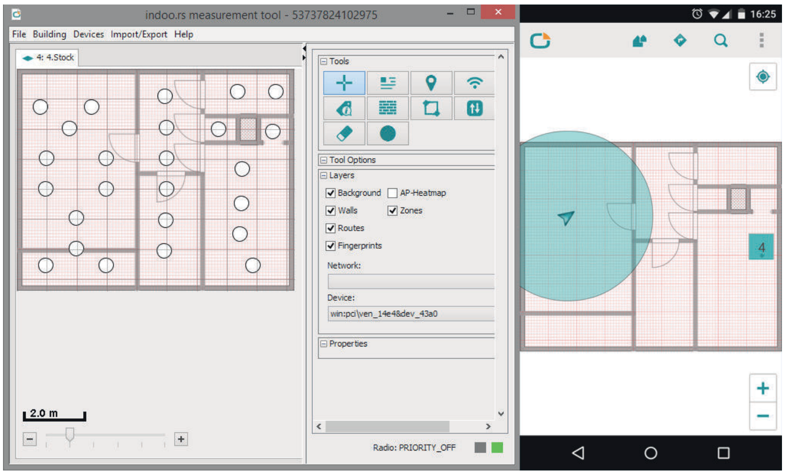
This helps to minimize errors due to human failure. A screenshot of each position is taken after each process in order to be able to later check the results in detail.

indoo.rs

Due to difficulties measuring position offsets with the provided Android application, a grid that has a line every 10 cm is used as an overlay over the actual plan. It is then loaded into indoo.rs’ proprietary Java-based configuration application (see figure 4.9), where the entirety of the preparation for the measuring phase can be done as well as the measurement for the training phase. The tool needs a conversion factor from the pixels of the floor plan image to a millimeter in real life. Changes will be saved in to the indoo.rs server so it can easily be accessed through the app.

The training phase requires the tester to carry a laptop to all the reference measurements and hold it still for about 15 to 20 seconds until the survey is finished. At every reference point, three surveys are automatically performed and the mean value is used. It is therefore very hard to create these measurements without human interference. For the reference measurements, the guidelines from the indoo.rs documentation [52] are met, namely:

- Measure about all three meters,
- Measure at least four or five points per room,
- Face away from the nearest wall when measuring,


indoo.rs measurement tool - 53737824102975
File Building Devices Import/Export Help
4: 4.Stock
Tools
Tool Options
Layers
Properties
2.0 m
16:25
4

☑ Background ☐ AP-Heatmap
☑ Walls ☑ Zones
☑ Routes
☑ Fingerprints

Network:

Device: win:pci\ven_14e4&dev_43a0

Radio: PRIORITY_OFF

Figure 4.9: On the left, the configuration tool is shown and on the right the simple client app can be seen.

- Spread the points evenly.

In the position phase, the same locations are used as in the experiment with SEASNIPS . The client device is an LG Nexus 5 with Android `5.0.1`. It will be placed at the respective locations and the tester will withdraw during a period of at least 13 seconds (that is, the time which SEASNIPS *high accuracy* positioning mode takes) to create similar conditions, as in the experiment with SEASNIPS . Since the app reacts to physical motion, i.e., the user tracking starts only when the device is moved, and the mobile phone will receive a firm "shake" to force a tracking attempt. After this amount of time, a screenshot of the device's display will be made before continuing to the next location. To measure the actual offsets, the screen shots will be used in combination with the screen shots of the previous experiment with SEASNIPS to create an overlay of the actual location. With an image processing tool, the pixel distance of these two can be measured. Now, because of the grid on the floor plan, which represents a physical distance, the pixel per meter can be calculated and offset in centimeters can be derived.

4.2.3 Setup

Experiment 1: Residential 65

The first target area is a roughly 65 m^2 flat on the fourth floor of a building from the 1950s. It features a room height of 2.48 m and the walls consist of bricks made from pressed debris. The floor plan can be examined in figure 4.10.

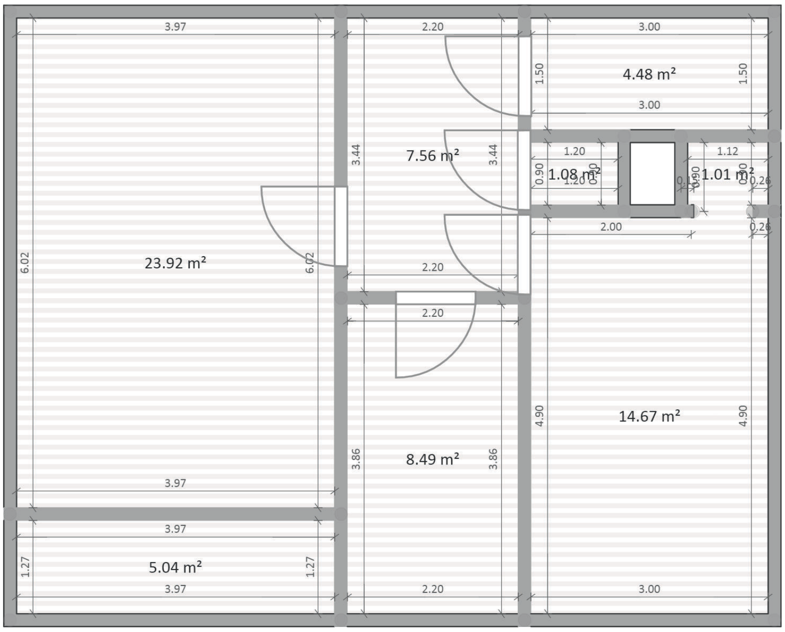


Figure 4.10: Floor plan of the flat. The outer measurements are 9.32 m × 7.45 m and the living area is exactly 64.16 m^2.

It is fully furnished and can be described as an average condo for two persons without any distinctive features apart from the balcony. As can be seen in figure 4.10, there is basically an AN in every room with about one or two rooms distance, and so the radio propagation should be fairly homogeneous among all of them. The base WLAN coverage is extensive, with about 20-30 APs in sight in all rooms of the target area, also making it perfectly suitable for a fingerprinting-based service.

Experiment 2: Studio 94

The second test area is located in a more remote industrial part of Vienna in a two-story building. The walls are made out of brick and concrete, and the premises are used as a studio for an artist. It has a very unusual composition, with one very long and large room, as can be seen in figure 4.12.

The main issue was - because of its secluded location - that there was practically no “natural” coverage of WLAN present, with merely five APs in range with very low readings of ≤ -85 dBm. To be able to create a suitable radio environment, a few APs where scattered semi-randomly over the area. Additionally, the AN used by SEASNIPS was left for indoo.rs to give it a fair try. The location of the said devices and the reference

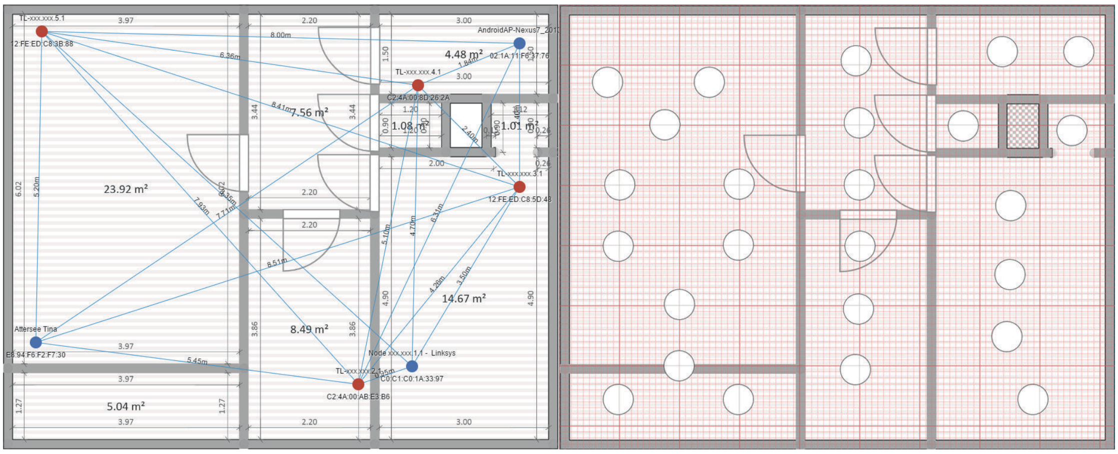


Figure 4.11: On the left, the locations of the ANs, in red, and ENs, in blue, can be seen. The right floor plan depicts all reference measure points as white circles for indoo.rs .

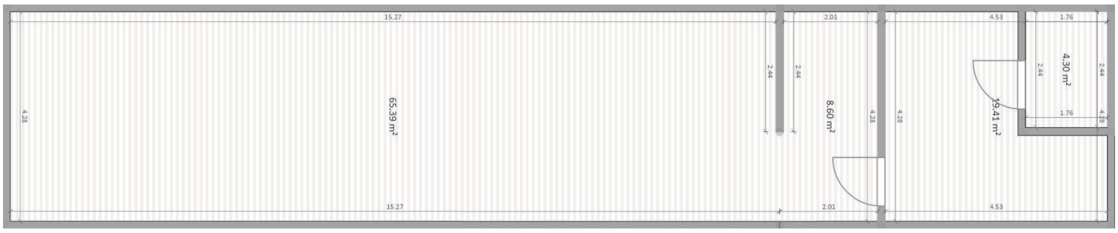


Figure 4.12: Floor plan of the studio. The outer measurements are 21.96 m × 4.44 m and the living area is exactly 93.5 m^2.

measure points can be compared with figure 4.13.

What sets it apart from the first experiment is that one large room is mainly used; therefore, line-of-sight (LOS)'s behavior is partly expected and it will be interesting to see how the two systems fare in this situation.

4.3 Results and Discussion

4.3.1 Experiment 1: Residential 65

Setup

Distributing the four ANs for Seasnips and measuring their positions took about 10 minutes. Setting up the system for the measuring phase took another 10 minutes. After about three hours, the system was evaluated and adjusted. This process required about 25 minutes, which makes it a grand total of 45 minutes of active work, whereby most of

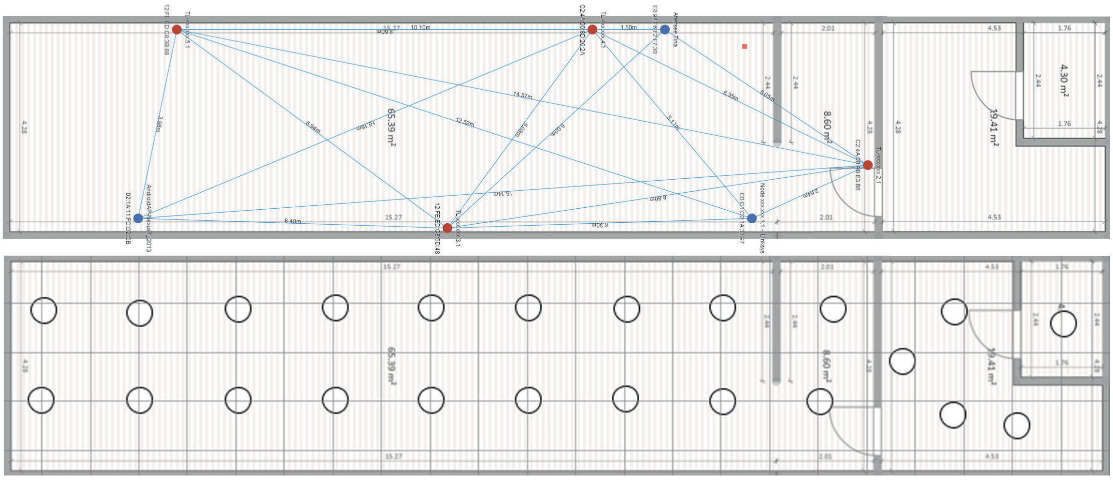


Figure 4.13: On the top, the locations of the ANs, in red, and ENs, in blue, can be seen. The bottom floor plan depicts all the reference measure points as white circles for indoo.rs .

it can be done sitting in front of a computer.

The process of setting indoo.rs up is the reverse to that of Seasnips . First, their proprietary tool was set up with the floor plan and then the measurements were begun. This is an extremely inconvenient process, because one has to hold the laptop, i.e., the measuring device, in their hands while simultaneously checking the correct location with the tape measure and clicking on the precise spot on the plan in the tool to start the survey and holding still during it. In addition, one has to be aware of his or her orientation towards the nearest walls to ensure the best results. A reference measurement took about $12 - 20$ sec to complete the survey. The setup is very quick, taking about 10 minutes, but the reference measurements of the 25 locations took about 30 minutes. It is therefore faster than Seasnips but more involved, since most of the time a laptop has to be balanced. The main advantage here is that there is no warm up time and the service can be used instantaneously after the measuring phase.

Trilateration Estimations

The initial trilateration estimation process for the ANs showed a partially correct result. Two out of four nodes where positioned nearly correctly - one was off by a couple of meters but in a correct relative position, while one was totally misplaced (*Node 4*). The reason for this was a non-correlating RSS value received from *Node 4*, as explained in more detail in the next paragraph. After adjusting the ANs to their actual positions, the locations of the ENs were estimated by Seasnips . Apart from one node, the output showed good quality. The Linksys AP was off by only about 2 m, and the TP-Link (`SSID: "Attersee"`) AP was broadly in the correct position. After correcting these offsets and deleting all EN where the actual position was unknown, the resulting map could be examined in the last step, as in figure 4.14.

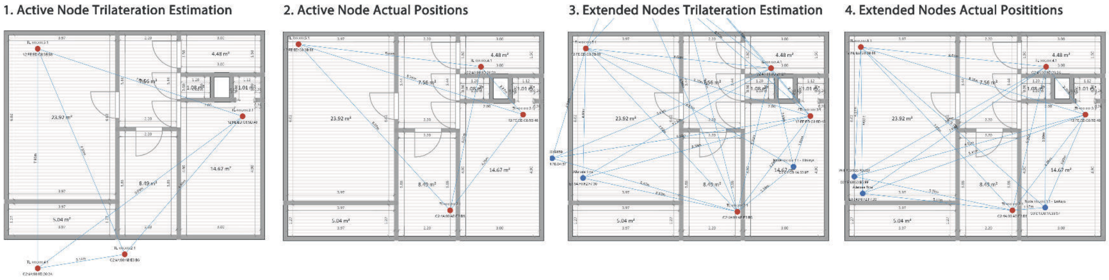
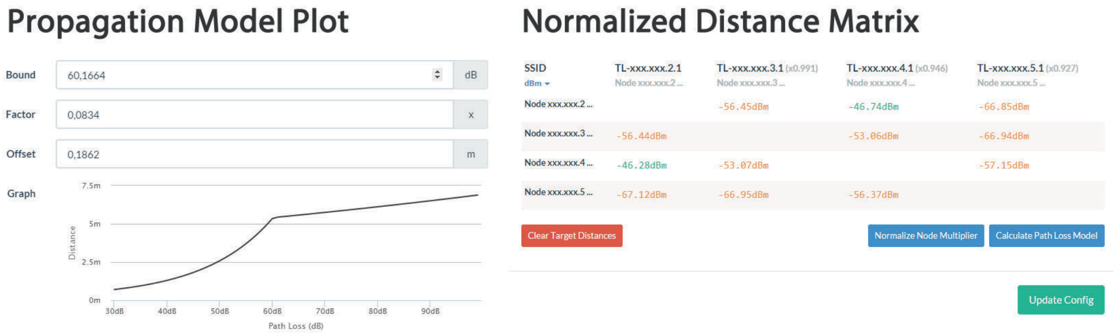

1. Active Node Trilateration Estimation
2. Active Node Actual Positions
3. Extended Nodes Trilateration Estimation
4. Extended Nodes Actual Posititions
23.92 m²
7.56 m²
4.48 m²
14.67 m²
8.49 m²
5.04 m²

Figure 4.14: The four steps of defining the positions of ANs and ENs.

Normalizing and Site Evaluation

The gain normalizer showed very good performance with a reduction in the error distance from 35.81 dBm to 3.07 dBm, which means that it created a nearly symmetric matrix, i.e., the best case scenario. The actual values can be compared in figure 4.15, where the RSS readings and the factors used for each AN are shown. After entering the actual distances, the propagation model can be adjusted. The default error distance was a decent 13.78 m, and could therefore only be enhanced by a small margin to 12.10 m. The resulting distance plot of the model can also be seen in figure 4.15. Since there are no further distances over 8 m, the automatic optimizer flattened the function from −60 dBm noticeably. This model cannot, of course, be applied to further distances, as mentioned before, but it works fine in the presented scenario.

The system had some trouble with the distance optimization because of how *Node 2* received its signal strength from *Node 3* and *Node 4*. Although the distance to *Node 4* is about 5.2 m over two rooms, and to *Node 3* it is 4.3 m over just one room, the RSS values do not reflect this, with −46.28 dBm and −56.44 dBm respectively. The system works best if the RSS and the distance correlate, but in cases like this where a smaller distance receives considerably less signal strength (by a factor of 10, converted to Watts), it is hard for Seasnips to adapt. This is also the reason why *Node 4* could not be correctly estimated in the trilateration process. Nevertheless, at least a slight optimization could be done. The comparable charts can be examined in 4.16.



Propagation Model Plot

Bound: 60,1664 dB

Factor: 0,0834 x

Offset: 0,1862 m

Normalized Distance Matrix

SSID dBm	TL-xxx.xxx.2.1 Node xxx.xxx.2 ...	TL-xxx.xxx.3.1 (x0.991) Node xxx.xxx.3 ...	TL-xxx.xxx.4.1 (x0.946) Node xxx.xxx.4 ...	TL-xxx.xxx.5.1 (x0.927) Node xxx.xxx.5 ...
Node xxx.xxx.2 ...		-56.45dBm	-46.74dBm	-66.85dBm
Node xxx.xxx.3 ...	-56.44dBm		-53.06dBm	-66.94dBm
Node xxx.xxx.4 ...	-46.28dBm	-53.07dBm		-57.15dBm
Node xxx.xxx.5 ...	-67.12dBm	-66.95dBm	-56.37dBm	

Figure 4.15: Plot of the used propagation model for Experiment 1 and the normalized distance matrix.

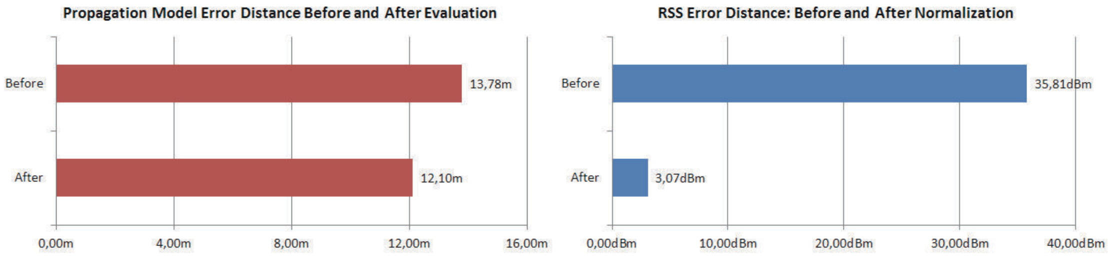


Figure 4.16: The achieved enhancements during the evaluation phase.

Accuracy

For the accuracy test, 14 location where chosen spread out over the whole area evenly, as can be seen in figure 4.17. The locations in the middle of the rooms but also near the wall, which usually are harder to estimate, were selected.

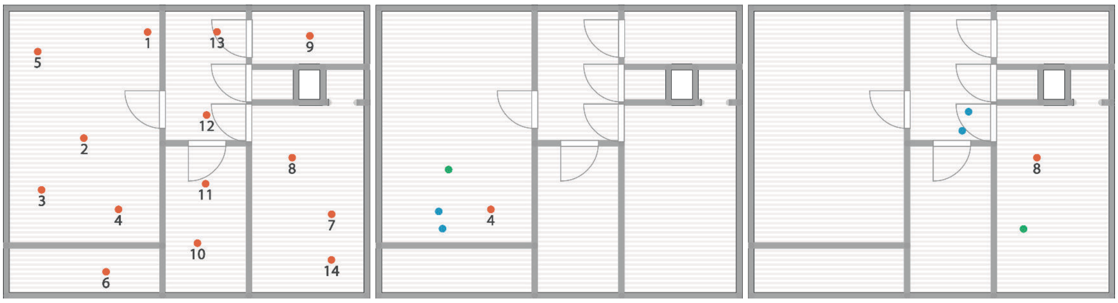


Figure 4.17: All the defined locations where the positioning was tested. On the right, two examples of the located positions are depicted, where the blue points represent the estimation from Seasnips and the green from indoo.rs .

In this setup, both systems fared quite well in the accuracy test, with a mean error of under 2 m for both SEASNIPS and indoo.rs , which provides practical usable results. What indoo.rs threw off in location test #9 was that it had problems detecting the position in the small non-windowed bathroom. The results can be examined in detail in figure 4.18 in combination with the location map in figure 4.17. All in all, indoo.rs showed a slightly better performance, also in one of the most important categories: room resolution. For the user, a system seems more unreliable if the estimated positions do not correlate with the room he currently is in, even if the offset is very low. Both systems showed good results here, but again indoo.rs outperformed SEASNIPS .

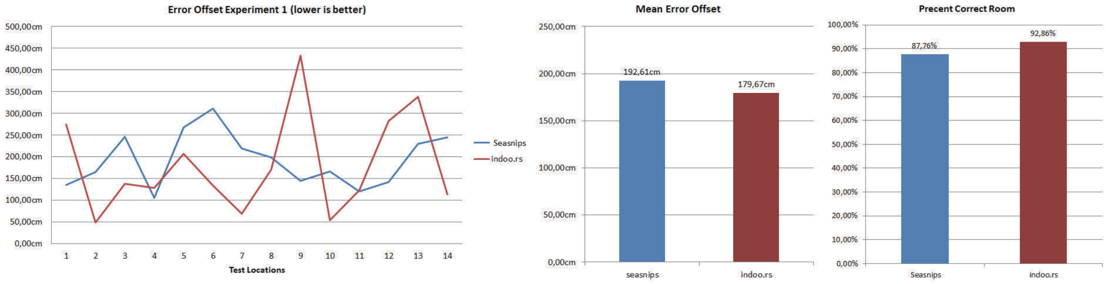


Figure 4.18: The results of the accuracy test over all 14 locations.

4.3.2 Experiment 2: Studio 94

Setup

The setup time for SEASNIPS was similar, with about one hour in total of active work and three hours of passive measurement phase. This time, though, indoo.rs took longer, since no laser tape measure was used but rather a regular one, making the process more involved and therefore needing about one hour and 20 minutes. Even though the setup looks more simple, containing the same count of reference measurement points as in Experiment 1, measuring in rooms over 15 meters in length with a regular tape is an intricate process.

Trilateration Estimations

In this scenario, the estimation clearly failed. The reason for this was the RSS readings from and to *Node 5*. From *Node 2*, signal strengths are received that follow the expected hierarchy of $RSS_{N4} > RSS_{N3}$, but for unknown reasons it also registers $RSS_{N5} > RSS_{N4}$ (although *Node 5* is 14.5 m away and *Node 4* only by 6.3 m). This is probably due to various effects of radio propagation (see section 2.2 “Radio Propagation: Large Scale

Models"), which are hard to predict. The details and the complete RSS distance matrix can be compared in figure 4.20.

Because of the problems mentioned above, the EN were also estimated incorrectly and had to be repositioned entirely, as seen in figure 4.19.

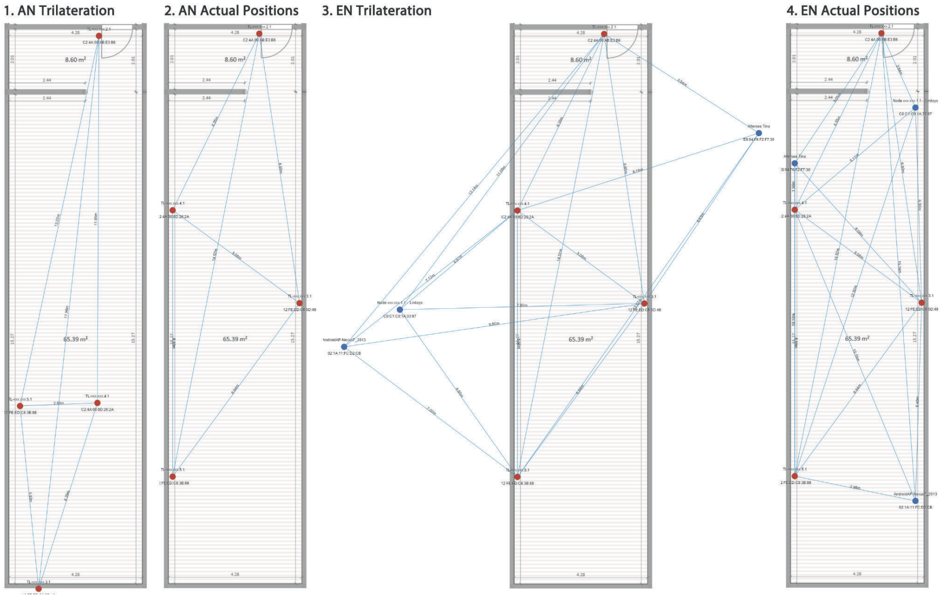
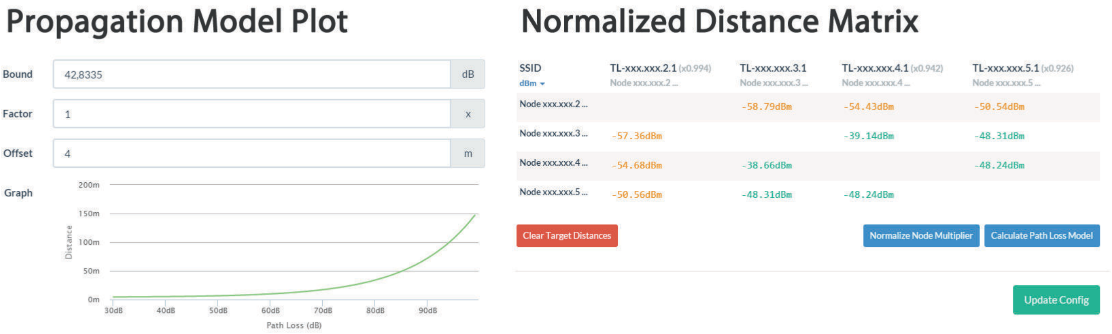


Figure 4.19: The four steps for defining the positions of ANs and ENs.

Normalizing and Site Evaluation

The gain normalizer again showed very good performance, with a reduction in the error offset to 4.40 dBm, creating a nearly symmetric matrix. The actual values can be compared in figure 4.20, where the RSS readings and the factors used for each AN are shown. Due to the non-correlating behavior of *Node 5*, as discussed above, the propagation model adjustment saw only average optimizations. The base error offset of 31.25 m could only achieve a reduction of about 6 m. This was only possible with manual adjustments, since the bounds of the automatic optimizer did not allow for the best results. Since the UI helps with a live error offset display, it was fortunately very easy to find a better configuration. First, the power loss coefficient of the underlying ITU indoor propagation model (see section 2.2.3 "ITU model for indoor attenuation") had to be set to one suitable for open spaces, which in this instance was the coefficient for commercial areas, in comparison to the default "office" setting, and as such the base offset had to be set to four meters. The plot of the resulting model can be seen in figure 4.20.



Propagation Model Plot

Bound: 42,8335 dB

Factor: 1 x

Offset: 4 m

Normalized Distance Matrix

SSID dBm ▾	TL-xxx.xxx.2.1 (x0.994) Node xxx.xxx.2 ...	TL-xxx.xxx.3.1 Node xxx.xxx.3 ...	TL-xxx.xxx.4.1 (x0.942) Node xxx.xxx.4 ...	TL-xxx.xxx.5.1 (x0.926) Node xxx.xxx.5 ...
Node xxx.xxx.2 ...		-58.79dBm	-54.43dBm	-50.54dBm
Node xxx.xxx.3 ...	-57.36dBm		-39.14dBm	-48.31dBm
Node xxx.xxx.4 ...	-54.68dBm	-38.66dBm		-48.24dBm
Node xxx.xxx.5 ...	-50.56dBm	-48.31dBm	-48.24dBm	

Figure 4.20: Plot of the used propagation model for Experiment 2 and the normalized distance matrix.

The comparison of the results after the evaluation can be seen in figure 4.21.

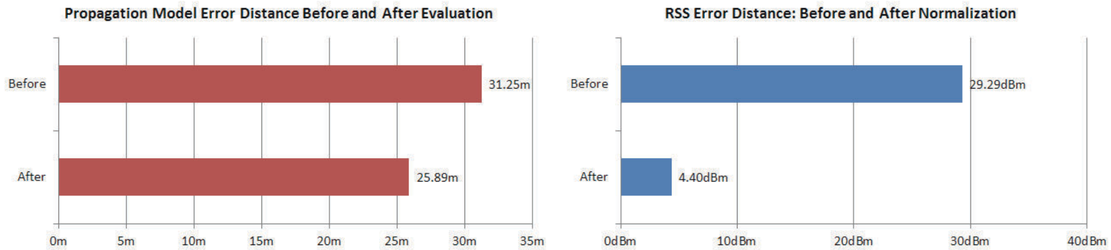


Figure 4.21: The enhancements achieved during the evaluation phase.

Accuracy

Because only four ANs were available for testing, only the one large room could be used as the target area, and so the test locations reflect this. Because of the lack of "natural" occurrences of APs, it would have needed one or two additional ANs to be able to include the top two rooms. All the selected test locations can be seen in figure 4.22.

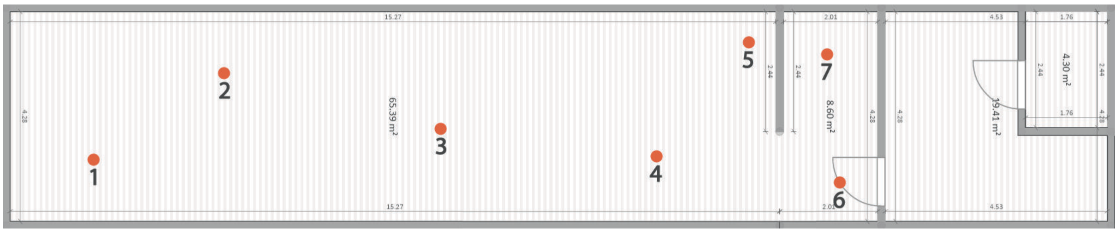


Figure 4.22: All the defined locations where the positioning was tested.

Both systems showed an average-to-bad performance in this test, although indoo.rs clearly comes out as the winner in this experiment, with a mean error offset of $\approx 2,6$ m against Seasnips with $\approx 3,6$ m (see fig 4.23 for details). Interesting to note is that, from test location one to five, both systems showed very similar performance but Seasnips clearly failed to estimate the position in the top part of the room, where the correlation error seems to affect the results significantly. The reason for this is the RSS readings from *Node 5*. It has to be said that some of the advantages of accuracy with indoo.rs are due to optimizations in the position client. For one, positions cannot - by definition - be out of the provided floor plan. The algorithm seems to choose the best position inside the floor here. While testing Seasnips , many estimations where beyond the borders of the floor plan because no bound check exists yet in the positing process.

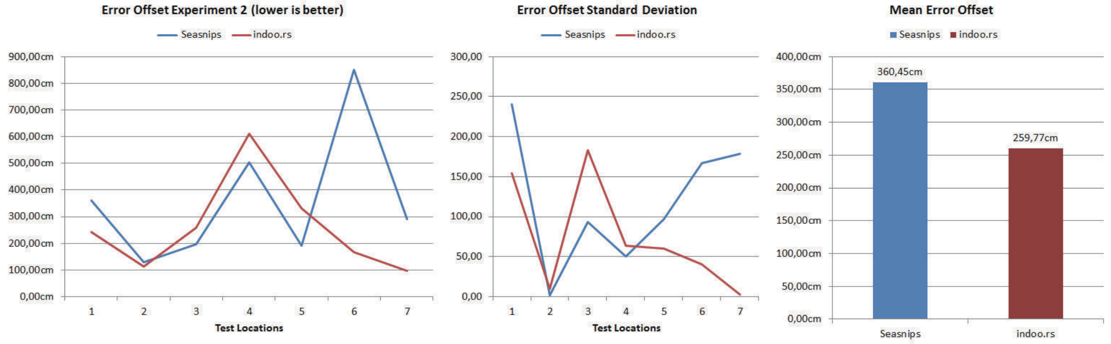


Figure 4.23: The results of the accuracy test over all seven locations. This includes the standard deviation of multiple measurements in a single location.

4.3.3 Summary of Results

In table 4.1, a brief summary of the findings from the performed experiment can be compared. It should be noted that the values in scalability increase linearly.

The main attributes of the system are summarized in the following list:

- adapts well to the specific environment
- good scalability
- low maintenance
- medium accuracy
- initial positioning with trilateration works well in most cases

	SEASNIPS	indoo.rs
Technology	WLAN RSS	WLAN RSS
Accuracy	2-3.5 m	2-2.5 m
Space Dim.	2D	2D
Cost	Low	Low
Scalability	180 min/300 m^2	389 min/300 m^2
Maintainability	continuous adaptation	full re-measurements

Table 4.1: Summary of the results.

- quality degrades with non-correlating RSS
- uses well-researched algorithms
- setup is easy and can be done from a single computer

CHAPTER 5

Conclusion and Outlook

5.1 Summary

In this thesis, contemporary methods for indoor positioning were researched and a novel approach based on received signal strength indicator (RSSI) with IEEE 802.11 which uses active sensor nodes was developed in order to enhance concepts currently used. This system's main goal is to provide a system architecture that enables easier maintenance, better scalability and robustness against continuous changes all the while keeping accuracy and precision comparable to the popular scene analysis (see section 2.1.2) technique. To validate this hypothesis, a fully functional prototype was implemented that could be used to evaluate this design.

However, before the results of the system will be discussed, first the basic question has to be considered: "Is received signal strength (RSS) with IEEE 802.11 a reliable and stable measurement?". The short answer is clearly: no. The main problem is not that some readings fluctuate considerably or that low signal strength sources are not discovered the majority of the time – the main problem is that some RSS do not at all correlate with physical distance. The correlation is not even expected to be linearly, but at least the following should always apply: higher distances should mean lower signal strengths. As seen in the previous experiments, this holds most of the time, but on a few occasions it does not and, depending upon the situation, this can only degrade the performance, either a little, as seen in Experiment 1, or else noticeably, as in Experiment 2. The differences in how non-correlating readings affect accuracy in these experiments are probably due to the line-of-sight (LOS)-type environment used in Experiment 2. The error weighs heavier because of the lower attenuation in free space, and therefore a greater difference in distance estimation. On the other hand, if the question becomes diffused to "Is RSS with IEEE 802.11 a usable method for rough distance estimation?", then the answer would be: yes. As the existence of indoor positioning and navigation services based received signal strength shows, it generally works. This could also be shown in the proposed system. It probably just needs techniques to counter these non-correlating instances, which is a subject of further research.

This is, unfortunately, also the main reason why the trilateration approach did not perform as expected. The initial positioning of the active nodes (ANs) were only partially useful. Often, at least one node was offset over a couple of meters. This, however, did not present a major problem, since the positions of the ANs are known and the correction is a matter of few mouse clicks. The positioning of the extended nodes (ENs) typically fared better and were within at least under two meters offset in most instances. However, since there was always at least one outlier, the result of the procedure could not be blindly trusted. As it seems that this coincides with the findings in the literature that trilateration with RSS does not present a sound approach for location estimation. The poor accuracy in this part of the design, however, did not impact upon the overall performance of the system considerably, since all these errors could be easily corrected manually.

Discussing one of the main performance metrics in an indoor positioning system - the accuracy - it generally can be said that *Site Evaluating Active Sensor Network Indoor Positioning System* (SEASNIPS) fared well against its fingerprinting counterpart. The mean accuracy was worse by only 6.72 % in Experiment 1 and 27.93 % in Experiment 2, as seen in figure 5.1. Due to the used service being commercially available for multiple years, it can be assumed that their positioning techniques saw a multitude of optimizations. Without these, the gap probably would be narrower.

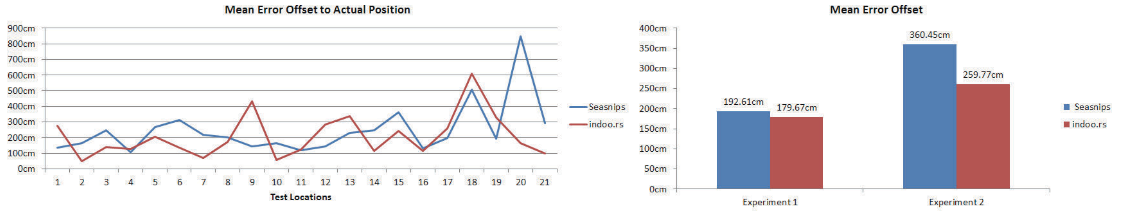


Figure 5.1: The aggregated results of all the experiments.

In addition, the precision was fairly good in Experiment 1, with a mean standard deviation of 26.33 and a more mixed performance of 118.33 in Experiment 2. As a comparison, indoo.rs showed a standard deviation of 73.59 in Experiment 2. It can therefore be said that, from this standpoint, Seasnips did prove its practical value in the field as an early prototype, but optimizations have to be found in LOS and non-correlating situations, as said before. It is expected that further optimizations will help overcome these flaws.

Now, the main advantage of Seasnips, as expected, is its superiority in terms of scalability and maintainability compared to indoo.rs . In the performed experiments, the setup duration is similar, but the effort is different. Meanwhile, with indoo.rs , all the measurements have to be done manually, with the proposed system most of the setup is done with a visual editor in front of the computer. Not only is the manual work taxing, it will also introduce errors due to human fault, skewing the fingerprinting map from this early phase. Moreover, for every re-measuring, the whole procedure has to be repeated. Though it seems coincidental that the duration of the setup is the same, after extrapolating the time it takes with an area four times as large, the effort for reference measurement for indoo.rs would just raise linearly, while with Seasnips it will not since most of the work is done on the system's user interface (UI).

In terms of cost, both systems require only a very small initial budget. Seasnips requires more hardware but only needs cheap consumer-grade devices, while indoo.rs costs more because of maintenance over time. Figure 5.2 shows the extrapolated time-effort for setting up the two different systems. The values are extrapolated from the performed tests and presume that, about every 15 m^2, one AN and about four-and-a-half reference measure points for scene analysis are needed. The setup time for a single AN and reference point is four minutes, and Seasnips needs another five minutes setup time per 15 m^2 while indoo.rs needs one.

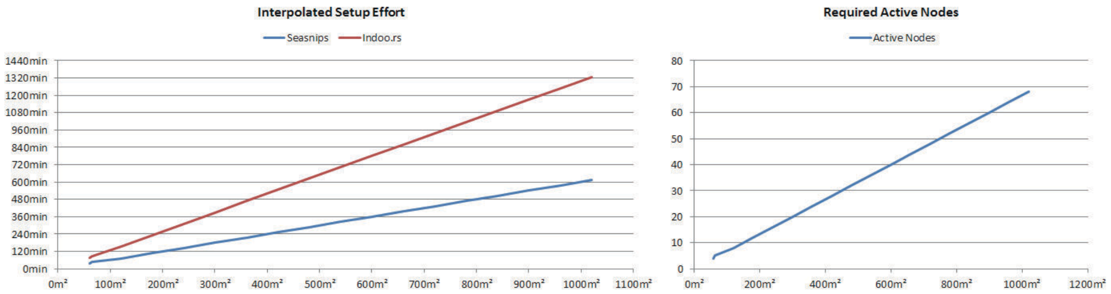


Figure 5.2: On the left is the extrapolated setup effort shown, and on the right the required AN are shown for different areas.

As such, while with indoo.rs every re-measurement needs the same effort as the first, it is drastically reduced with Seasnips since only the computer model has to be adjusted.

No specific bias emerged as to what the best appliance for the proposed system was from the experiments; therefore, it can be considered as a general purpose system where the mean error offset of several meters is acceptable. It was clearly shown that a LOS could present a hostile environment, but this cannot be generally derived from the performed tests. As such, all-in-all, it showed a promising performance similar to the fingerprinting approach, and with several advantages in other areas, namely scalability and maintainability.

5.2 Future Work

As it currently stands, Seasnips provides a complete implementation for creating a position service for arbitrary environments. However, as already stated in chapter 3 "Design and Implementation", some cutbacks had to be made to align efforts with this work. In particular, there are some issues that need to be addressed first in order to truly make it practically relevant. Furthermore, these will involve testing and adapting for human interference and user tracking.

Human interference is a complicated matter and there is no easy solution to counter this. Most systems simply acknowledge the fact that this issue degrades the quality of the service, although Bahl et al. [4] tries to address it with multiple signal maps, each for a different conical orientation. Since this is not possible to do with the architecture, Seasnips presents another approach that needs to be contemplated. On option would be to facilitate an orientation-aware sensor in the clients (usually smart phones featuring acceleration sensors and a gyroscope) and then use an attenuation model to estimate the probable impact of absorption with the help of the radio map as simple geometric calculation (under the assumption that the user holds the device approximately in front of him all the time).

The other important problem is user tracking, since often such services are not only used for discrete positioning but also for navigation. Here, an optimized Nearest Neighbor (NN) algorithm might be a feasible option to apply. As explained in Equation 2.1.2 "Nearest Neighbor", a version similar to that discussed, *constrained search-space nearest neighbor*, which keeps the user's previous position history, might help in finding positions that are physically possible to reach, drastically minimizing the error rate. A similar concept was also applied to Radar [5]. This also has to be verified with extensive testing.

5.2.1 General Enhancements

Though propagation models work fairly well in most cases, they have difficulties in others. To encounter this, a more sophisticated solution is needed, namely ray tracing (see section 2.2.1). This could help improve the quality of the estimations drastically, and probably would make propagation models obsolete as the main source for attenuation estimation. This would, however, involve major changes in the architecture, since precise floor plans with additional meta-information, like floor- and wall-types, are required. A similar approach to that proposed by the ARIADNE [56] project by Ji et al. seems feasible, since it generalizes radio wave properties that contribute little to the overall signal strength but which will complicate the system and demand more computational power.

Building on this, expanding the mapped area to the third dimension could help the accuracy, especially with using ray tracing. For this, floor attenuation has to be considered as well as the placement heights of ANs, further adding complexity to the configuration for more precise estimations. As previously stated, it has to show if this change would positively affect the overall performance of the service or make it more susceptible to configurations that are exaggerated in the multiple steps of mapping.

5.2.2 Advanced Site Adaptation

The presented adaptation components significantly increase the quality of service, although some improvements could help to make the system more adaptable and stable. The first is to use a more sophisticated adaptable propagation model. The current model merely offers a basic offset and factor, and a definition of a single "joint" as seen as a function plot. This could be addressed either with additional joints supporting different gains for different ranges of path loss or using a completely new model that only generally reflects the wave propagation with multiple easy customizable parameters. This, however, requires a deeper understanding of the mathematical basis of propagation models, which was not within the scope of this work.

As observed by different researchers [4, 58], RSS fluctuates throughout the day because of the radio wave absorption of persons present in continuously changing environments (e.g., closed/opened doors, etc.). In most cases, these variations are systematic, where, for instance, throughout the day the RSS values are lower compared to the evening where they increase due to less interference. This could be implemented with time-dependent radio maps. Currently, a signal strength map is valid for the whole day, but if the system would allow us to create time slots, different measurements for various times only affect maps responsible for this time slot, and the positing service would also only select the currently active map.

When mapping large areas, sometimes it is not possible to distribute ANs evenly and the sections only receive weak signal strengths, lowering the quality of the service. Another problem involves spaces with a high count of obstacles drastically changing the wave propagation. This could be adjusted by some manual measuring in these areas and incorporating them into the *signal strength map* to positively influence the validity in some regions.

5.2.3 Algorithmic Enhancements

SEASNIPS uses only Nearest Neighbor (NN) at the moment. With the high-density grid approach, this makes sense since the return on investment of implementing more complicated algorithms seems dim. However, without comparative testing, this is hard to say for certain. It is probably unrealistic to expect a significant increase in overall precision and accuracy, but enhancements in certain special cases may be achieved. Interesting candidates would include Smallest M-vertex Polygon Algorithm (SMP), as described in section 2.1.2 “Smallest M-Vertex Polygon”, or the clustering approach proposed by Ji et al. with their system ARIADNE [56]. Of course, if time is not an issue, a full comparison with a Support Vector Machine (SVM) and a neural network might show whether these methods justify the high implementation cost. The best solution is probably a combination of approaches or else using multiple techniques conditionally for different scenarios.

In another area, filters, like the one proposed by Teuber et al. [107], could improve the quality of the RSS measurements, especially in cases where it is not possible or feasible to gather sample sizes of several hundreds. Viable candidates would be the aforementioned fuzzy logic filter or a Kalman Filter, for instance. Moreover, as already stated, it has to be re-evaluated as to which statistical property or which combination is best-used in this case.

5.2.4 Active Client

To complete SEASNIPS , a capable active client has to be implemented. The currently-used passive clients (see section 3.1.2 “Positioning Client”) suffice for testing, though they are not ideal for actual user positioning. The target platforms are smart phones with either *Android*, *iOS* or *Windows Phone*, although limitations have to be considered, for example in *iOS* where, without additional tempering, no access to Wireless Local Area Network (WLAN) signal level measurements can be obtained. The client should feature a fluid, high-detail map where the current position is indicated with additional features, like context information depending upon the intended benefits of the service for the user. This would also need extended features from the server so that maps and additional information can be acquired. Detailed maps are usually implemented as multi-level tiles where maps of different levels of detail are tiled so that they can be zoomed transparently into each other without exceeding the memory budget and sacrificing performance. This is a very efficient approach, although it adds complexity and requires considerable effort to implement and maintain.

Since, as stated before, modern smartphones contain additional sensors, it has to be examined which of those available can be used to enhance the position estimations. Possible options include an orientation-awareness sensors and pedometer for better user tracking, for instance, or a barometer for height estimations (which have become a more common form of sensor in recent years).

Glossary

802.11a defined in 1999 this wireless network standard has a maximum raw data rate of 54 Mbit/s in the 5 Ghz band. 34

802.11ac is an amendment to IEEE 802.11, published in December 2013, that builds on 802.11n. Changes compared to 802.11n include wider channels in the 5 GHz band, more spatial streams, and the addition of Multi-user MIMO. 34, 35

802.11b defined in 1999 this wireless network standard has a maximum raw data rate of 11 Mbit/s and uses the same media access method defined in the original standard. 34, 35, 40, 41, 87

802.11g is an amendment from 2003 to the IEEE 802.11 specification that extended throughput to up to 54 Mbit/s using the same 2.4 GHz band as 802.11b. 34, 35, 87

802.11n is a wireless networking standard that uses multiple antennas to increase data rates in comparison to 802.11g. 34, 35, 87

access point is similar to Wireless LAN routers with the difference that they do not necessarily have Ethernet ports. x, 93, *see* Wireless Local Area Network

AngularJS is an open-source web application framework maintained by Google and a community of individual developers and corporations. 62, 63

Apache CouchDB is an open source database that focuses on ease of use. 62, 63

Apache License is a free software license written by the Apache Software Foundation (ASF) that allows the user of the software the freedom to use the software for any purpose, to distribute it, to modify it, and to distribute modified versions of the software, under the terms of the license, without concern for royalties. 63

Assisted Global Positioning System is a location services that combines GPS with TDOA trilateration in cellular networks. 10, 93, *see* Global Positioning System

base64 is a encoding that represents binary data in an ASCII string format. 51

basic access authentication is a simple form of authentication specificed in Hypertext Transfer Protocol (HTTP) that grants access if the correct user name and password is provided with the request header. 51

Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the band of 2.4 GHz). 2, 30, 32, 35, 93

CDDL Common Development and Distribution License is a free software license, produced by Sun Microsystems, based on the Mozilla Public License (MPL). 63

centroid is the geometric center of a two-dimensional polygon which is the arithmetic mean position of all the points in the shape. 16

CRON is a time-based job scheduler in Unix-like computer operating systems. 64, 67

Decibel-isotropic is measure of isotropic antenna gain which uniformly distributes energy in all directions; most common used antenna in access point (AP). 33, 93

distance matrix is a matrix containing the distances, taken pairwise, of a set of points. 53–56

effective radiated power is a standardized theoretical measurement of radio frequency (RF) energy. 26, 94

Euclidean distance The Euclidean distance is the "ordinary" distance between two points in Euclidean space. 14, 19, 40, 42

feedforward In neural networks, feedforward is an artificial neural network where connections between the units do not form a directed cycle. This is different from recurrent neural networks. 17

Fraunhofer region is another term for far-field which is a region of a electromagnetic field around an object. 22

Global Positioning System is like GLONASS and Galileo a space-based satellite navigation system that provides location information. 1, 94

IEEE 802.11 is a set of specifications for implementing wireless local area network (WLAN) computer communication. 1, 27–30, 34, 36, 38–40, 46, 49, 50, 87, 92

Intel Core i5-2520M was a fast high-end dual core processor for mobile devices at the time of introduction in Q1 2011 ; the base clock speed is 2.5 GHz but due to Turbo Boost it can reach 3 GHz (2 cores active) and 3.2 GHz (1 core active). 65, 67

Java is a general-purpose computer object-oriented programming language. 18, 24, 62, 63, 69

JavaScript is a dynamic computer programming language. 62, 63

k nearest neighbour used as algorithm in fingerprinting for finding the position of a mobile user. 15, 40, 95

Kalman Filter Kalman filtering, also known as linear quadratic estimation (LQE), is an algorithm that uses a series of measurements observed over time, containing noise and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. 10, 75

Linux is a POSIX-complaint computer operation system under the model of free and open-source software. 50, 59, 60

Lua is a lightweight multi-paradigm programming language designed as a scripting language with extensible semantics as a primary goal. 60, 61

Mahalanobis distance The Mahalanobis distance is a measure of the distance between a point P and a distribution D. 15

Manhattan distance The Manhattan distance between two points is the sum of the absolute differences of their Cartesian coordinates, similar to car in the island of Manhattan which takes the shortest path between two intersections. 14, 40, 53, 55, 65, 69, *see* Euclidean distance

MapReduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster. 62

midpoint circle algorithm is an algorithm used to determine the points needed for drawing a circle on a raster. 56

MIT License is a free software license originating at the Massachusetts Institute of Technology (MIT) which permits reuse within proprietary software provided all copies of the licensed software include a copy of the MIT License terms and the copyright notice. 63

mode is the value that appears most often in a set of data in distribution. 37

Multi-Layer Perceptron is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs; used in neural networks with the fingerprinting approach. 17, 95

NoSQL often interpreted as “Not Only SQL” it provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases. 62

OpenWRT is an operating system / embedded operating system based on the Linux kernel, and primarily used on embedded devices to route network traffic. 59, 60

outlier in statistics, an outlier is an observation point that is distant from other observations. 64

path loss is defined as the difference in dB between the effective transmitted power and the received power. 22

PHP is a server-side scripting language designed for web development. 60

Quartz is a job scheduling library that can be integrated into a wide variety of Java applications. 62–64

Radial Basis Function is a real-valued function whose value depends only on the distance from the origin. 17, 95

radio-frequency identification RFID is the wireless use of electromagnetic fields to transfer data, for the purposes of automatically identifying and tracking tags attached to objects. 19, 96

Rayleigh distribution is a continuous probability distribution for positive-valued random variables. 28

received signal strength is the power present in a received radio signal. ix, 19, 31, 64, 66, 96, *see* received signal strength indicator

received signal strength indicator is a metric for radio signal receivers usually measured in dBm a logarithmic scale for milliwatt [49, p. 41]. 1, 31, 96

Round Trip Time is the length of time it takes for a signal to be sent plus the length of time it takes for an acknowledgment of that signal to be received. 10, 96

Servlet is a Java programming language program that extends the capabilities of a server. 62, 63

shell script is a computer program designed to be run by the Unix shell; dialects of shell scripts are considered to be scripting languages. 50

simulated annealing is a generic probabilistic metaheuristic for the global optimization problem of locating a good approximation to the global optimum of a given function in a large search space. 43

Smallest M-vertex Polygon Algorithm Algorithm used among other thing as an position estimation in the fingerprinting approach. 16, 96

Sparse Matrix it is a data-type that is usually memory-optimized for matrices that are filled only sparse. 69

Support Vector Machine is a supervised learning model with associated learning algorithms that analyze data and recognize patterns which is a research topic in combination with scene analysis. 17, 18, 96

time difference of arrival similar to time of arrival but measures the time difference of a signals from the emitter at three or more synchronized receiver sites. 6, 9, 96, *see* time of arrival

time of arrival is an acronym for time of arrival which is the travel time of a radio signal from a single transmitter to a remote single receiver. 2, 96

torus is a donut-shaped object, much like a o-ring. 33

tranmission power output is the actual amount of power of radio frequency energy that a transmitter produces at its output. 32, 96

Trie a trie is an ordered tree data structure that is used to store a dynamic set or associative array where the keys are usually strings. 69

UTF-8 is a character encoding capable of encoding all possible characters in Unicode. 35

Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states. 41

WaveLAN was a brand name for a family of wireless networking technology sold by NCR, AT&T, and Lucent introduced in 1988; it was generally not compliant with IEEE 802.11 although compatible families existed. 2, 40

weighted k nearest neighbor used as algorithm in fingerprinting for finding the position of a mobile user. 42, 97, *see* k nearest neighbour

Wi-Fi is the trademark name for WLANs using 2.4 or 5 Ghz radio waves with the IEEE 802.11 standard. It is a play on the audiophile term Hi-Fi. 30, *see* Wireless Local Area Network

Wireless Local Area Network is a wireless computer network that links multiple clients; nowadays mostly based on the IEEE 802.11 standard. 1, 97, *see* IEEE 802.11

Acronyms

A-GPS Assisted Global Positioning System. 10, 82, *see* Assisted Global Positioning System

AES Advanced Encryption Standard. 34

AKH Wiener Allgemeine Krankenhaus. 4

AN active node. x, 46, 49–54, 56–59, 61, 63, 65, 66, 68, 69, 74, 75

AoA Angle of Arrival. 10, 82

AP access point. x, 1, 3, 11, 14, 19, 33–40, 46, 49, 50, 57, 64, 68, 88, *see* access point

API application programming interface. 2, 46, 51, 62

ASF Apache Software Foundation. 87

BSS basic service set. 35, 36, 94

BSSID basic service set identifier. 35, 50, 59, 69

BT Bluetooth. 2, 19, 82, *see* Bluetooth

CDF cumulative distribution function. 64

CERN Conseil Européen pour la Recherche Nucléaire. 62

CGI Common Gateway Interface. 60

CSS Cascading Style Sheets. 62

dB decibel. 22–24, 26, 27, 31

dBi Decibel-isotropic. 33, *see* Decibel-isotropic

dBm decibel-milliwatts. 31–33, 50, 58, 66, 69

DECT Digital Enhanced Cordless Telecommunications. 28, 82

DOM Document Object Model. 63

EiRP Equivalent isotropically radiated power. 32

EN extended node. 46, 57, 58, 68, 69

ERP effective radiated power. 26, 32, *see* effective radiated power

ESS extended service set. 36

ESSID extended service set identifier. 36

ETSI European Telecommunications Standards Institute. 35

EURO-COST European Co-operative for Scientific and technical research. 26

FAF Floor Attenuation Factor. 41

FSPL free-space path loss. 22, 26

Ghz gigahertz. 26

GPL GNU General Public License. 63

GPS Global Positioning System. 1, 2, 5, 9, 10, *see* Global Positioning System

GSM Global System for Mobile Communications. 26

HTTP Hypertext Transfer Protocol. 50, 51, 59, 60, 62, 88

HTTPS Hypertext Transfer Protocol Secure. 52

IBSS independent BSS. 36

IR infrared radiation. 19, 82

ISM Industry, Scientific and Medical. 34, 35

ITU International Telecommunication Union. 29

JAX-RS Java API for RESTful Web Services. 62, 63

JDK Java Development Kit. 69

JSON JavaScript Object Notation. 50, 62, 63

KNN k nearest neighbour. 15, 16, 18, 19, 40, 43, 59, 82, *see* k nearest neighbour

KWNN k weighted nearest neighbor. 15–18

LAN local area network. 3, 50

LOS line-of-sight. 20–22, 26, 27, 39, 54

MAC media access control. 35, 38

MB megabyte. 60

MCV model-view-controller. 62

Mhz megahertz. 29

MIMO multiple-input and multiple-output. 34

MIT Massachusetts Institute of Technology. 90

MLP Multi-Layer Perceptron. 17, 82, *see* Multi-Layer Perceptron

MPL Mozilla Public License. 88

mW milliwatt. 31, 32

NIC network interface controller. 3, 36, 64

NLOS non-line-of-sight. 27, 37

NN Nearest Neighbor. 16, 19, 40, 41, 47, 59, 69, 73, 75

NNSS nearest neighbor in signal space. 40

RaPSor Radio Propagation Simulator. 24, 25

RBF Radial Basis Function. 17, *see* Radial Basis Function

REST Representational State Transfer. 50, 59, 62

RFID radio-frequency identification. 19, 41–43, 82, *see* radio-frequency identification

RSS received signal strength. ix, x, 3, 6, 8, 9, 11, 12, 14, 16, 17, 19, 20, 30, 31, 36–40, 42, 46, 49, 50, 53–55, 58, 59, 61, 64–66, 68, 69, 74, 75, *see* received signal strength

RSSI received signal strength indicator. 1, 8, 11, 18, 19, 30–32, 36–41, 43, 46, 82, *see* received signal strength indicator

RTT Round Trip Time. 10, *see* Round Trip Time

SCS service control server. ix, 46, 49, 50, 52, 59, 63, 64

Seasnips *Site Evaluating Active Sensor Network Indoor Positioning System*. ix, 45–47, 54–57, 73, 75

SMP Smallest M-vertex Polygon Algorithm. 16, 75, 82, *see* Smallest M-vertex Polygon Algorithm

SNR signal-to-noise ratio. 49

SSID service set identifier. 35, 36, 50, 51, 59

SSL Secure Socket Layer. 52

SVM Support Vector Machine. 17, 75, 82, *see* Support Vector Machine

T-R transmitter-receiver. 20, 22, 24, 26, 29, 41

TDOA time difference of arrival. 6, 9, 10, 18, 82, *see* time difference of arrival

TLS Transport Layer Security. 52, 60

TOA time of arrival. 2, 6, 9, 10, 18, *see* time of arrival

TPO tranmission power output. 32, 33, *see* tranmission power output

U-TDOA uplink time difference of arrival. 9, *see* time difference of arrival

UI user interface. x, 62–64, 66, 68

VLAN virtual local area network. 35, 51

WAF Wall Attenuation Factor. 41

WAN wide area network. 50, 52

WKNN weighted k nearest neighbor. 42, 43, 82, *see* weighted k nearest neighbor

WLAN Wireless Local Area Network. 1–3, 5, 19, 26, 30, 32, 33, 36, 39, 40, 43, 46, 47, 49–51, 53, 57, 59, 60, 76, 82, 84, 85, *see* Wireless Local Area Network

WPA Wi-Fi Protected Access. 34

WPA2 Wi-Fi Protected Access II. 34

XML Extensible Markup Language. 50

Bibliography

[1] D. Akerberg. Properties of a TDMA Pico Cellular Office Communication System. In *Vehicular Technology Conference, 1989, IEEE 39th*, pages 186–191 vol.1, May 1989.

[2] Neil Ashby. Relativity and the Global Positioning System. *Physics Today*, 55(5):41–47, 2002.

[3] Lilian Aveneau, Emilie Masson, and Pierre Combeau. RaPSor: a Radio Propagation Simulator, Presentation and Use Cases. Product presentation, 2010.

[4] P. Bahl and V.N. Padmanabhan. RADAR: An In-building RF-based User Location and Tracking System. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784 vol.2, 2000.

[5] Paramvir Bahl, Venkata N Padmanabhan, and Anand Balachandran. Enhancements to the RADAR User Location and Tracking System. Technical report, Microsoft Research, 2000.

[6] Leslie W Barclay. *Propagation of Radiowaves*, volume 502. Iet, 3rd edition, 2012. ISBN: 1849195781.

[7] Joe Bardwell. Converting Signal Strength Percentage to dBm Values. `https://madwifi-project.org/attachment/wiki/UserDocs/RSSI/Converting_Signal_Strength.pdf`, 2002 (accessed December 7, 2014). WildPackets.

[8] K. Benkic, M. Malajner, P. Planinsic, and Z. Cucej. Using RSSI Value for Distance Estimation in Wireless Sensor Networks based on ZigBee. In *Systems, Signals and Image Processing, 2008. IWSSIP 2008. 15th International Conference on*, pages 303–306, June 2008.

[9] R.C. Bernhardt. Macroscopic Diversity in Frequency Reuse Radio Systems. *Selected Areas in Communications, IEEE Journal on*, 5(5):862–870, Jun 1987.

[10] Henry L. Bertoni, W. Honcharenko, L.R. Macel, and H. Xia.

[11] Chatschik Bisdikian, Ibm Corporation, and Owned The Bluetooth. An Overview of the Bluetooth Wireless Technology. *IEEE Communications Magazine*, 39:86–94, 2001.

[12] Azzedine Boukerche, Horacio ABF Oliveira, Eduardo F Nakamura, and Antonio AF Loureiro. Localization Systems for Wireless Sensor Networks. *IEEE Wireless Communications*, 14(6):6–12, 2007.

[13] Mauro Brunato and Roberto Battiti. Statistical Learning Theory for Location Fingerprinting in Wireless {LANs} . *Computer Networks*, 47(6):825 – 845, 2005.

[14] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less Low-cost Outdoor Localization for very Small Devices. *Personal Communications, IEEE*, 7(5):28–34, Oct 2000.

[15] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[16] Y. Chapre, P. Mohapatra, S. Jha, and A. Seneviratne. In *Local Computer Networks (LCN), 2013 IEEE 38th Conference on*.

[17] Yongguang Chen and Hisashi Kobayashi. Signal Strength based Indoor Geolocation. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 1, pages 436–439, 2002.

[18] T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos. Empirical Calculation of Shadowing Deviation for Complex Indoor Propagation Topologies at 2.4 GHz. In *Ultra Modern Telecommunications Workshops, 2009. ICUMT '09. International Conference on*, pages 1–6, Oct 2009.

[19] T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos. Site-specific Validation of ITU indoor Path Loss Model at 2.4 GHz. In *World of Wireless, Mobile and Multimedia Networks Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, pages 1–6, June 2009.

[20] Marc Ciurana, Francisco Barceló, and Sebastiano Cugno. Indoor Tracking in WLAN Location with TOA Measurements. In *Proceedings of the 4th ACM International Workshop on Mobility Management and Wireless Access*, MobiWac '06, pages 121–125, New York, NY, USA, 2006. ACM.

[21] D.C. Cox, R.R. Murray, and A.W. Norris. 800-MHz Attenuation Measured in and around Suburban Houses. *AT T Bell Laboratories Technical Journal*, 63(6):921–954, July 1984.

[22] B.P. Crow, I. Widjaja, Jeong Geun Kim, and P.T. Sakai. IEEE 802.11 Wireless Local Area Networks. *Communications Magazine, IEEE*, 35(9):116–126, Sep 1997.

[23] Eraldo Damosso and Luis M Correia. *COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report.* European Commission, 1999.

[24] RealTimeLocationSystem Ecived. Indoor Positioning System Manufacturer. `https://www.flickr.com/photos/realtimelocationsystem/7364539118/in/photostream/`, 2012 (accessed January 15th, 2015).

[25] Christian Esposito and Massimo Ficco. Deployment of RSS-Based Indoor Positioning Systems. *International Journal of Wireless Information Networks*, 18(4):224–242, 2011.

[26] ETSI. EN 301.893 BRAN; Broadband Radio Access Networks (BRAN); 5GHz high Performance RLAN; v.1.7.0. *HARMONIZED EUROPEAN STANDARD*, 2012.

[27] A. Falsafi, K. Pahlavan, and Ganning Yang. Transmission Techniques for Radio LAN's-a Comparative Performance Evaluation using Ray Tracing. *Selected Areas in Communications, IEEE Journal on*, 14(3):477–491, Apr 1996.

[28] Shih-Hau Fang and Tsung-Nan Lin. Indoor Location System Based on Discriminant-Adaptive Neural Network in IEEE 802.11 Environments. *Neural Networks, IEEE Transactions on*, 19(11):1973–1978, Nov 2008.

[29] FCC. FCC Part 15.247 Rules Systems Using Digital Modulation . *Radio Frequency Device: operation with the bands*, pages 2400–2483, 2009.

[30] E. Ferro and F. Potorti. Bluetooth and Wi-Fi Wireless Protocols: a Survey and a Comparison. *Wireless Communications, IEEE*, 12(1):12–26, Feb 2005.

[31] I. Forkel and M. Lott. A Multi-Wall-and-Floor Model for Indoor Radio Propagation. In *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, volume 1, pages 464–468 vol.1, 2001.

[32] Harald T Friis. A note on a Simple Transmission Formula. *proc. IRE*, 34(5):254–256, 1946.

[33] Michael Gauthier. 2.4 GHz Wi-Fi Channels (802.11b,g WLAN). `https://commons.wikimedia.org/wiki/File:2.4_GHz_Wi-Fi_channels_(802.11b,g_WLAN).svg`, 2009 (accessed December 8, 2014).

[34] I.A. Getting. Perspective/Navigation - The Global Positioning System. *Spectrum, IEEE*, 30(12):36–38, 1993.

[35] M. Ghaddar, L. Talbi, and T.A. Denidni. Human Body Modelling for Prediction of Effect of People on Indoor Propagation Channel. *Electronics Letters*, 40(25):1592–1594, Dec 2004.

[36] D.B. Green and A.S. Obaidat. An Accurate Line of Sight Propagation Performance Model for Ad-Hoc 802.11 Wireless LAN (WLAN) Devices. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 5, pages 3424–3428 vol.5, 2002.

[37] U. Grossmann, M. Schauch, and S. Hakobyan. RSSI based WLAN Indoor Positioning with Personal Digital Assistants. In *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2007. IDAACS 2007. 4th IEEE Workshop on*, pages 653–656, Sept 2007.

[38] Youngjune Gwon, R. Jain, and T. Kawahara. Robust Indoor Location Estimation of Stationary and Mobile Users. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1032–1043 vol.2, March 2004.

[39] Y.R. Hamdy and S.A. Mawjoud. Performance Assessment of U-TDOA and A-GPS Positioning Methods. In *Future Communication Networks (ICFCN), 2012 International Conference on*, pages 99–104, April 2012.

[40] M. Hata. Empirical Formula for Propagation Loss in Land Mobile Radio Services. *Vehicular Technology, IEEE Transactions on*, 29(3):317–325, Aug 1980.

[41] Conrad E Heidenreich. Measures of Distance Employed on 17th and early 18th Century Maps of Canada. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 12(2):121–137, 1975.

[42] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger, and James Collins. Global Positioning System. Theory and Practice. *Global Positioning System. Theory and practice., by Hofmann-Wellenhof, B.; Lichtenegger, H.; Collins, J.. Springer, Wien (Austria), 1993, 347 p., ISBN 3-211-82477-4, Price DM 79.00. ISBN 0-387-82477-4 (USA).*, 1, 1993.

[43] Ville Honkavirta, Tommi Perala, Simo Ali-Loytty, and Robert Piché. A Comparative Survey of WLAN Location Fingerprinting Methods. In *Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on*, pages 243–251. IEEE, 2009.

[44] IEEE. IEEE Standard Definitions of Terms for Antennas. *IEEE Std 145-1983*, pages 1–31, June 1983.

[45] IEEE. IEEE Standards for Local and Metropolitan Area Networks: Overview and Architecture. *IEEE Std 802-1990*, 1990.

[46] IEEE. IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks-Specific Requirements Part Ii: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11g-2003 (Amendment to IEEE Std 802.11, 1999 Edn. (Reaff 2003) as amended by IEEE Stds 802.11a-1999, 802.11b-1999, 802.11b-1999/Cor 1-2001, and 802.11d-2001)*, pages i–67, 2003.

[47] IEEE. IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks-Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Medium Access Control (MAC) Security Enhancements. *IEEE Std 802.11i-2004*, pages 1–175, 2004.

[48] IEEE. IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput. *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pages 1–565, Oct 2009.

[49] IEEE. IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pages 1–2793, March 2012.

[50] IEEE. IEEE Standard for Information technology– Telecommunications and information exchange between systemsLocal and metropolitan area networks– Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications–Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz. *IEEE Std 802.11ac-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, and IEEE Std 802.11ad-2012)*, pages 1–425, Dec 2013.

[51] European Cooperation in the Field of Scientific and EURO-COST 231 Technical Research. Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands. *COST 231 TD (91) 73. Rev 2*, 1991.

[52] indoo.rs. indoo.rs documentation.

[53] informability.com. Image Credits for WRT54GL image. `https://store.informability.com/store/images/u_04090033.jpg`, accessed Jannuary 9, 2015.

[54] F. Izquierdo, M. Ciurana, F. Barcelo, J. Paradells, and Enrica Zola. Performance Evaluation of a TOA-based Trilateration Method to Locate Terminals in WLAN. In *Wireless Pervasive Computing, 2006 1st International Symposium on*, pages 1–6, Jan 2006.

[55] Nam-Ryul Jeon, Chang-Hoon Lee, Noh-Gyoung Kang, and Seong-Cheol Kim. Performance of Channel Prediction Using 3D Ray-tracing Scheme Compared to Conventional 2D Scheme. In *Communications, 2006. APCC '06. Asia-Pacific Conference on*, pages 1–6, Aug 2006.

[56] Yiming Ji, Saâd Biaz, Santosh Pandey, and Prathima Agrawal. ARIADNE: A Dynamic Indoor Signal Map Construction and Localization System. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services*, MobiSys '06, pages 151–164, New York, NY, USA, 2006. ACM.

[57] Seokmin Jung and Woontack Woo. UbiTrack: Infrared-based User Tracking System for Indoor Environment. *International Conferece on Artificial Reality and Telexisitence (ICAT04)*, pages 1345–1278, 2004.

[58] K. Kaemarungsi and P. Krishnamurthy. Properties of indoor received signal strength for WLAN location fingerprinting. In *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, pages 14–23, Aug 2004.

[59] Kamol Kaemarungsi and Prashant Krishnamurthy. Modeling of Indoor Positioning Systems based on Location Fingerprinting. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1012–1022. IEEE, 2004.

[60] M. Kanaan and K. Pahlavan. A Comparison of Wireless Geolocation Algorithms in the Indoor Environment. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, volume 1, pages 177–182 Vol.1, March 2004.

[61] P. Karlsson, C. Bergljung, E. Thomsen, and H. Borjeson. Wideband Measurement and Analysis of Penetration Loss in the 5 GHz band. In *Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th*, volume 4, pages 2323–2328 vol.4, 1999.

[62] J.M.P. Keenan and A.J. Motley. Personal Communication Radio Coverage in Buildings at 900 MHz and 1700 MHz. *Electronics Letters*, 24(12):763–764, Jun 1988.

[63] J.M.P. Keenan and A.J. Motley. Radio Coverage in Buildings. *British telecom technology Journal*, 8(1):19–24, 1990.

[64] Mikkel Baun Kjærgaard, Georg Treu, Peter Ruppel, and Axel Küpper. Efficient Indoor Proximity and Separation Detection for Location Fingerprinting. In *Proceedings of the 1st International Conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications*, MOBILWARE '08, pages 1:1–1:8, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[65] P. Kontkanen, P. Myllymaki, T. Roos, H. Tirri, K. Valtonen, and H. Wettig. Topics in probabilistic location estimation in wireless networks. In *Personal, Indoor*

and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on, volume 2, pages 1052–1056 Vol.2, Sept 2004.

[66] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Lydia E. Kavraki, and Dan S. Wallach. Robotics-based Location Sensing Using Wireless Ethernet. *Wirel. Netw.*, 11(1-2):189–204, January 2005.

[67] William Lee. *Mobile communications design fundamentals.* John Wiley&Sons, New York, 1993.

[68] Charles E Leiserson, Ronald L Rivest, Clifford Stein, and Thomas H Cormen. *Introduction to Algorithms.* MIT press, 2001.

[69] Binghao Li, James Salter, Andrew G Dempster, and Chris Rizos. Indoor Positioning Techniques Based on Wireless LAN. In *LAN, First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, pages 13–16. Citeseer, 2006.

[70] Binghao Li, Yufei Wang, Hyung Keun Lee, Andrew Dempster, and Chris Rizos. Method for Yielding a Database of Location Fingerprints in WLAN. *IEE Proceedings-Communications*, 152(5):580–586, 2005.

[71] Christophe Liebe, Pierre Combeau, Alain Gaugue, Yannis Pousset, Lilian Aveneau, Rodolphe Vauzelle, and Jean-Marc Ogier. Ultra-wideband Indoor Channel Modelling using Ray-Tracing Software for through-the-wall Imaging Radar. *International Journal of Antennas and Propagation*, 2010, 2010.

[72] Tsung-Nan Lin and Po-Chiang Lin. Performance Comparison of Indoor Positioning Techniques based on Location Fingerprinting in Wireless Networks. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 1569–1574 vol.2, June 2005.

[73] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of Wireless Indoor Positioning Techniques and Systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, 2007.

[74] Wen-Chung Liu. Optimal Design of Dualband CPW-fed G-shaped Monopole Antenna for WLAN Application. *Progress In Electromagnetics Research*, 74:21–38, 2007.

[75] Matthias Lott. On the Performance of an Advanced 3D Ray Tracing Method. In *In Proceedings of European Wireless & ITG Mobile Communication*, 1999.

[76] G. Lui, T. Gallagher, Binghao Li, A.G. Dempster, and C. Rizos. Differences in rssi readings made by different wi-fi chipsets: A limitation of wlan localization. In *Localization and GNSS (ICL-GNSS), 2011 International Conference on*.

[77] Eladio Martin, Oriol Vinyals, Gerald Friedland, and Ruzena Bajcsy. Precise Indoor Localization Using Smart Phones. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 787–790, New York, NY, USA, 2010. ACM.

[78] Paul M Maxim, Suranga Hettiarachchi, William M Spears, Diana F Spears, Jerry Hamann, Thomas Kunkel, and Caleb Speiser. Trilateration Localization for Multi-Robot Teams. *ICINCO-RA*, 2:301–307, 2008.

[79] Microsoft. WLAN_ASSOCIATION_ATTRIBUTES Structure (Windows). `http://msdn.microsoft.com/en-us/library/windows/desktop/ms706828%28v=vs.85%29.aspx`, 2010 (accessed December 7, 2014).

[80] F. Naya, H. Noma, R. Ohmura, and K. Kogure. Bluetooth-based Indoor Proximity Sensing for Nursing Context Awareness. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pages 212–213, Oct 2005.

[81] Chahé Nerguizian, Charles Despins, and Sofiene Affes. Indoor geolocation with received signal strength fingerprinting technique and neural networks. In *Telecommunications and Networking-ICT 2004*, pages 866–875. Springer, 2004.

[82] LionelM. Ni, Yunhao Liu, YiuCho Lau, and AbhishekP. Patil. LANDMARC: Indoor Location Sensing Using Active RFID. *Wireless Networks*, 10(6):701–710, 2004.

[83] D. Niculescu and B. Nath. Ad-hoc Positioning System (APS) using AOA. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1734–1743 vol.3, March 2003.

[84] Yoshihisa Okumura, Eiji Ohmori, Tomihiko Kawano, and Kaneharu Fukuda. Field Strength and its Variability in VHF and UHF land-mobile Radio Service. *Rev. Elec. Commun. Lab*, 16(9):825–73, 1968.

[85] OpenWrt Project. openwrt.

[86] Dhruv Pandya, Ravi Jain, and E. Lupu. Indoor Location Estimation using Multiple Wireless Technologies. In *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, volume 3, pages 2208–2212 vol.3, Sept 2003.

[87] Ambili Thottam Parameswaran, Mohammad Iftekhar Husain, Shambhu Upadhyaya, et al. Is Rssi a Reliable Parameter in Sensor Localization Algorithms: An Experimental Study. *Field Failure Data Analysis Workshop*, F2DA09, 2009.

[88] Rick Pemble. Coax Attenuation Chart - W4RP. `http://www.w4rp.com/ref/coax.html`, 2005 (accessed December 8, 2014).

[89] Phongsak Prasithsangaree, Prashant Krishnamurthy, and Panos K Chrysanthis. On Indoor Position Location with Wireless LANs. In *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, volume 2, pages 720–724. IEEE, 2002.

[90] OpenWrt Project. OpenWrt Web Presence. `http://openwrt.org`, accessed Jannuary 9, 2015.

[91] Michael Quan, Eduardo Navarro, and Benjamin Peuker. Wi-Fi Localization Using RSSI Fingerprinting.

[92] ITU Radiocommunication. Recommendation ITU-R P. 1238-7: Propagation Data and Prediction Models for the Planning of Indoor Radiocommunication Systems and Radio Local Area Networks in the Frequency Range 900 MHz to 100 GHz, 2012.

[93] B. Raman and K. Chebrolu. Experiences in using WiFi for Rural Internet in India. *Communications Magazine, IEEE*, 45(1):104–110, Jan 2007.

[94] T. S. Rappaport. The Wireless Revolution. *Comm. Mag.*, 29(11):52, 61–71, Nov 1991.

[95] Theodore S Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, 2nd edition, 2002.

[96] Frank Reichenbach, Alexander Born, Dirk Timmermann, and Ralf Bill. A Distributed Linear Least Squares Method for Precise Localization with Low Complexity in Wireless Sensor Networks. In PhillipB. Gibbons, Tarek Abdelzaher, James

Aspnes, and Ramesh Rao, editors, *Distributed Computing in Sensor Systems*, volume 4026 of *Lecture Notes in Computer Science*, pages 514–528. Springer Berlin Heidelberg, 2006.

[97] A. Roxin, J. Gaber, M. Wack, and A. Nait-Sidi-Moh. Survey of Wireless Geolocation Techniques. In *Globecom Workshops, 2007 IEEE*, pages 1–9, Nov 2007.

[98] Andreas Savvides, Chih-Chieh Han, and Mani B Strivastava. Dynamic Fine-Grained Localization in Ad-Hoc Networks of Sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 166–179. ACM, 2001.

[99] M. Saxena, P. Gupta, and B.N. Jain. Experimental Analysis of RSSI-based Location Estimation in Wireless Sensor Networks. In *Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on*, pages 503–510, Jan 2008.

[100] S.Y. Seidel and T.S. Rappaport. 914 MHz Path Loss Prediction Models for Indoor Wireless Communications in Multifloored Buildings. *Antennas and Propagation, IEEE Transactions on*, 40(2):207–217, Feb 1992.

[101] John S Seybold. *Introduction to RF propagation*. John Wiley & Sons, 2005.

[102] Navin Kumar Sharma. A Weighted Center of Mass Based Trilateration Approach for Locating Wireless Devices in Indoor Environment. In *Proceedings of the 4th ACM International Workshop on Mobility Management and Wireless Access*, MobiWac '06, pages 112–115, New York, NY, USA, 2006. ACM.

[103] B. Sklar. Rayleigh fading Channels in Mobile Digital Communication Systems .I. Characterization. *Communications Magazine, IEEE*, 35(7):90–100, Jul 1997.

[104] Asim Smailagic, Jason Small, and Daniel P Siewiorek. Determining User Location for Context Aware Computing through the use of a Wireless LAN Infrastructure. *Institute for Complex Engineered Systems Carnegie Mellon University, Pittsburgh, PA*, 15213, 2000.

[105] Warren L. Stutzman and Thiele A. Gary. *Antenna Theory and Design*. Wiley Online Library, 3rd edition, 2013. ISBN 978-0-470-57664-9.

[106] N. Swangmuang and P. Krishnamurthy. Location Fingerprint Analyses Toward Efficient Indoor Positioning. In *Pervasive Computing and Communications, 2008.*

PerCom 2008. Sixth Annual IEEE International Conference on, pages 100–109, March 2008.

[107] Andreas Teuber, Bernd Eissfeller, and Thomas Pany. A Two-Stage Fuzzy Logic Approach for Wireless LAN Indoor Positioning. In *Proc. IEEE/ION Position Location Navigat. Symp*, volume 4, pages 730–738, 2006.

[108] TP-LINK. Image Credits for TP-LINK images. `http://www.tp-link.com/resources/images/products/large/`, accessed Jannuary 9, 2015.

[109] Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, 1993.

[110] wien.gv.at. Neues Allgemeines Krankenhaus. `https://www.wien.gv.at/wiki/index.php/Neues_Allgemeines_Krankenhaus`, 2015 (accessed March 25th, 2015).

[111] The Hong Kong Polytechnic University Wireless Innovative Navigation System. Wireless Tracking Analysis on Wi-Fi Signal Strength. `http://www.comp.polyu.edu.hk/lab/wins/research_area.htm`, 2008 (accessed January 20th, 2015).

[112] Chao-Lin Wu, Li-Chen Fu, and Feng-Li Lian. WLAN location determination in e-home via support vector classification. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 2, pages 1026–1031 Vol.2, 2004.

[113] W.M. Yeung and J.K. Ng. An Enhanced Wireless LAN Positioning Algorithm based on the Fingerprint Approach. In *TENCON 2006. 2006 IEEE Region 10 Conference*, pages 1–4, Nov 2006.

[114] Moustafa A Youssef, Ashok Agrawala, and A Udaya Shankar. WLAN Location Determination via Clustering and Probability Distributions. In *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 143–150. IEEE, 2003.

[115] Zengbin Zhang, Xia Zhou, Weile Zhang, Yuanyang Zhang, Gang Wang, Ben Y. Zhao, and Haitao Zheng. I Am the Antenna: Accurate Outdoor AP Location Using Smartphones. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 109–120, New York, NY, USA, 2011. ACM.

[116] Yilin Zhao. Standardization of Mobile Phone Positioning for 3G Systems. *Communications Magazine, IEEE*, 40(7):108–116, Jul 2002.

APPENDIX A

Propagation Reference Tables

Environment	Path Loss Exponent n
Free space	2
Urban cellular radio	2.7 to 3.5
Shadowed urban cellular radio	3 to 5
In building line-of-sight	1.6 to 1.8
Obstructed in building	4 to 6
Obstructed in factories	2 to 3

Table A.1: Path loss exponents for the Log-distance path loss model [95, p.104]

Frequency	Residential	Office	Commercial
900 Mhz		33	20
1.2 to 1.3 Ghz		32	22
1.8 to 2 Ghz	28	30	
2.4 Ghz	28	30	
3.5 Ghz		27	
4 Ghz		28	22
5.2 Ghz (apartment)	30	31	
5.2 Ghz (house)	28		
5.8 Ghz		24	
60 Ghz		22	17
70 Ghz		22	

Table A.2: Power loss coefficients, N, for indoor transmission loss calculation [92, p.4]

Frequency	Type	Residential	Office	Commercial
900 Mhz	1 floor		9	
900 Mhz	2 floors		19	
900 Mhz	3 floors		24	
1.8 to 2 Ghz		$4n$	$15 + 4(n-1)$	$6 + 3(n-1)$
2.4 Ghz	Apartment	10	14	
2.4 Ghz	House	5		
3.5 Ghz	1 floor		18	
3.5 Ghz	2 floors		26	
5.2 Ghz	Apartment	13	16 (1 floor)	
5.2 Ghz	House	7		
5.8 Ghz	1 floor		22	
5.8 Ghz	2 floors		28	

Table A.3: Floor penetration loss factors with n being the number of floors penetrated, for indoor transmission loss calculation ($n \geq 1$) [92, p.5]

Office Building 1	FAF (dB)	σ
1 Floor	12.9	7.0
2 Floors	18.7	2.8
3 Floors	24.4	1.7
4 Floors	27.0	1.5
Office Building 2		
1 Floor	16.2	2.9
2 Floors	27.5	5.4
3 Floors	31.6	7.2
SF PacBell		
1 Floor	13.2	9.2
2 Floors	18.1	8.0
3 Floors	24.0	5.6
4 Floors	27.0	6.8
5 Floors	27.1	6.3

Table A.4: The average floor attenuation factor in dB as measured by Seidl [95, 100]

APPENDIX B

API

B.1 Active Node

`ping`

Description

A simple requests requiring hardly any computational power therefore can be called multiple times in a small time-frame. Meant as first step for node discovery and to continuously check if the device is online and functioning correctly.

Url Structure

`http://192.168.x.x/cgi-bin/ping`

Method

`GET`

Query Parameters

none

Returns

`200` if everything is ok or either a timeout or `500` if a problem occurred

`iwinfo`

Description

The main service responsible for provide hardware informations and returning details of a WLAN survey.

Url Structure

`http://192.168.x.x/cgi-bin/iwinfo`

Method

`GET`

Query Parameters

none

adapter the adpater's name is expected here as query parameter (e.g. `url?adapter=wlan0`)

Returns

A list of all adapters and additional hardware informations if no parameter is given and the result of a survey and additional hardware informations for the adapter given in the query string

Example XML

```
http://192.168.x.x/cgi-bin/iwinfo?adapter=wlan0
```

```
<?xml version="1.0" encoding="UTF-8"?>
<router>
  <adapter name="wlan0">
    <info><![CDATA[ ... ]]></info>
    <scan><![CDATA[ ... ]]></scan>
    <txpowerlist><![CDATA[ ... ]]></txpowerlist>
    <assoclist><![CDATA[ ... ]]></assoclist>
  </adapter>
  <adapter-list> <![CDATA[ ... ]]></adapter-list>
  <date> <![CDATA[ ... ]]></date>
  <uptime> <![CDATA[ ... ]]></uptime>
```

```
    <ifconfig> <![CDATA[ ... ]]></ifconfig>
    <cpuinfo> <![CDATA[ ... ]]></cpuinfo>
    <meminfo> <![CDATA[ ... ]]></meminfo>
    <version> <![CDATA[ ... ]]></version>
</router>
```

`reboot`

Description

A simple service that starts a device reboot. This is a maintenance task to be able to fix simple software malfunction remotely like e.g. the web server of a node does not work correctly.

Url Structure

`http://192.168.x.x/cgi-bin/reboot`

Method

`GET`

Query Parameters

none

Returns

`200` if everything is ok or either a timeout or `500` if a problem occurred

B.2 Server

`getPossibleNetworksForGivenSurvey`

Description

Returns one or more probable networks i.e. possible target areas when there are multiple floors mapped for instance for the given WLAN survey.

Url Structure

`http://server-url/positioning/networks`

Method

`POST`

Query Parameters

freq the used frequency, possible values: `WLAN_2_4Ghz`, `WLAN_5Ghz`

Payload

```
{
   "wlanSSID1": {
      "ssid":"wlanSSID1",
      "macAddress":"00:11:22:33:44:55",
      "channel":6,
      "date":"yyyy-MMM-dd HH:mm:ss",
      "signalStrengths":[35.5,36,35]
   },
   "wlanSSID2": {
      ...
   }
}
```

Returns

A list of `SensorNetworks`

Example JSON

```
{
   [
      {"networkId":"1234567890abcd",
      "networkName":"Network1",
      "description":"...",
      "createDate":"yyyy-MMM-dd HH:mm:ss",
      ...}
   ]
}
```

`getProbablePositions`

Description

Returns the positions determined by the algorithm with the highest probability in the given `SensorNetwork`.

Url Structure

`http://server-url/positioning/position`

Method

`POST`

Query Parameters

freq the used frequency, possible values: `WLAN_2_4Ghz`, `WLAN_5Ghz`

networkId the id of the `SensorNetwork`

Payload

```
{
    "wlanSSID1": {
        "ssid":"wlanSSID1",
        "macAddress":"00:11:22:33:44:55",
        "channel":6,
        "date":"yyyy-MMM-dd HH:mm:ss",
        "signalStrengths":[35.5,36,35]
    },
    "wlanSSID2": {
        ...
    }
}
```

Returns

A list of calculated positions relative to the grid of the *signal strength map*

Example JSON

```
{
   [
      {"x":13,
      "y":21,
      "tileLengthCm":10,
      "probabilityValue":0.8}
   ]
}
```