Universidade de Aveiro
2006

Departamento de Electrónica e Telecomunicações

PEDRO MIGUEL FERREIRA CLARO

SISTEMA DE POSICIONAMENTO LOCAL BASEADO EM REDES WI-FI

Universidade de Aveiro
2006

Departamento de Electrónica e Telecomunicações

PEDRO MIGUEL FERREIRA CLARO

SISTEMA DE POSICIONAMENTO LOCAL BASEADO EM REDES WI-FI

LOCAL POSITIONING SYSTEM BASED ON WI-FI NETWORKS

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica do Dr. Nuno Borges de Carvalho, Professor associado do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro

“I much prefer the sharpest criticism of a single intelligent man to the thoughtless approval of the masses.”

Johannes Kepler

o júri

presidente

Prof. Dr. Nuno Miguel Gonçalves Borges de Carvalho
professor associado do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

Prof. Dr. Carlos Lisboa Bento
professor associado do Departamento de Informática da Universidade de Coimbra

Prof. Dr. Rui Valadas
professor associado do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro

agradecimentos

Agradeço ao meu orientador, Prof. Nuno Borges, pelas ideias, sugestões, críticas e objectividade que sempre estiveram presentes durante este trabalho.
Agradeço também aos meus colegas no Instituto de Telecomunicações pela camaradagem e apoio.
Por últimos uma palavra especial para a minha família que me recordou nos momentos difíceis, o que é mais importante.

A todos um obrigado,

Pedro Claro

palavras-chave

Localização, redes sem fio, propagação, redes neuronais, Wi-Fi, posicionamento

resumo

Este trabalho pretende descrever na totalidade um sistema de localização em ambientes interiores, desde a sua criação, desenvolvimento e implementação. Este sistema de localização tem a capacidade de inferir a posição de dispositivos sem fios utilizando uma simples solução por software. Ao contrário de muitos sistemas de localização actuais, não necessita de hardware adicional.

Na primeira parte é apresentada uma introdução geral e uma descrição dos sistemas e técnicas de localização actuais. No resto do trabalho são descritos os vários passos realizados de modo a atingir a solução e sistema finais. Na parte final são apresentados os resultados alcançados bem como uma análise de possíveis melhoramentos.

keywords

Location, wireless networks, propagation , neural networks, Wi-Fi, position

abstract

This work describes a complete indoor location system, from its creation, development and deployment. This location system is a capable way of retrieving the position of wireless devices using a simple software solution, no additional hardware is necessary.

In the first part it, a general introduction is made and a description of current location systems and techniques is presented. One the second and following sections a description of the several steps performed to achieve a final solution are presented. The last sections describe the results achieved and an analysis of possible future work is presented.

Table of contents

Table of figures

Table list

1 Introduction

The location paradigm began years ago when systems like Decca, OMEGA, Alpha and Loran C were developed. Loran C was developed by the United States Navy during World War II. This system main objective was to help US and UK military ships navigation. Several radio beacon towers were deployed along sea coast. Using these radio beacons and their known location, the ships were capable of locating themselves. This was an important tactical advantage and it is still used today, although some modifications were made. Current location systems are more technical advanced, but almost all of them use the same principle of Loran – radio triangulation. This is the case of the well known Global Positioning System (GPS). GPS uses geo-stationary satellites as radio beacons, and provides almost a global coverage real time location system. Triangulation is a simple concept to understand. As radio waves have a constant and well known speed, knowing the

location of the radio beacon source, one can calculate easily the distance traveled by it as long as the transmission time is also known.
The importance of a location system can be defined using scopes of application. Keeping a real track of military units was the first common application of a location system. But the location paradigm can be applied to provide a new set of services – marketing, work, daily life, emergency situations, traveling, fleet management, etc.

GPS was first developed to be an aid to US military forces as a tactical asset capable of giving a crucial advantage in a war scenario. A few years ago, the USA granted limited public access to its system allowing the development of new location applications to the general public. There was a boom in GPS receivers' sales allowing the general public to benefit from a global location system. Cars are being sold with navigation capabilities, mobile phones became an aid to track people in emergency situations (ie: E911), fleets are being managed real-time, etc. Satellite navigation besides its huge coverage has the clear disadvantage of not providing indoor coverage. In fact, getting a clear signal in narrow streets is difficult or almost impossible. Indoor location can be important when one thinks of services that can be applied in huge buildings, like shopping centers, office buildings, museums, warehouses, universities, etc. Imagine a warehouse equipped with a location system, one can know exactly the position of a package allowing an increase in the company's efficiency. Think of a museum where tourists can receive in their cell phones important information depending on the place they are or the art work they are seeing. Imagine receiving on your cell phone great price discounts when you walk along a store in your favorite shopping center. This one might not be so great, but the applications and advantages of an accurate and reliable indoor location system are tremendous. This is the aim of this work, to develop a low cost, reliable and accurate location system using today's technology.

Radio frequency signals are present everywhere and almost at any time. From GSM mobile phones to Bluetooth devices, the variety of frequencies and technology is enormous. The goal of this work was to develop a location system capable of taking advantage of this availability of signals. Techniques such as Time of Arrival (ToA), Time Difference of Arrival (TDoA) or Angle of Arrival (AoA) [1] require a special infrastructure and special hardware. Deploying a location system capable of competing with existing services requires a low cost, accurate and scalable solution.

This work takes in account that there is a well defined telecommunication infrastructure, where mobile devices roam freely. The tracking of these devices is possible using a technique, called radio-frequency fingerprinting. This technique takes advantage of the received signal strength intensity (RSSI) measured by the mobile devices. The RSSI values are collected in respect to every base station that can be detected. Using these values, and after some calculations and comparisons, it is possible to acquire the mobile device current position. The source of signals can be GSM, WiFi, FM stations, TV stations, etc. However, as a proof of concept, a full location system applied to wireless LANs was constructed and fully tested. This location system was deployed in different indoor scenarios, where it proved to be an accurate, low-cost and easy to install solution.

2 Current state of the art

Current location systems can be separated into two different types depending on using dedicated hardware or just being software based. The type of information that a location system can provide also differs. Some systems provide symbolic information, like a room or a division. Sometimes this kind of information might be enough depending on the needs of the application. Others try to pinpoint an exact location described by the usage of spatial coordinates.

In the first type of location systems, satellite navigation systems, like North American GPS or the future European Galileo, are mainly focused on providing position for outdoor environments. These systems provide a global coverage with a three to five meters average error available for public usage. Although satellite navigation systems have good accuracy they are not suitable to indoor environments where a good clear view to the sky is not available. Other systems are mainly focused on indoor location environments. Systems like Active-Bat [2] developed by AT&T and Cricket [3] use ultrasound time of flight measurements, others like Active-Badge [4] use small infrared tags that provide symbolic information, like a name of a room. Infrared has limited range hence it has not become very popular. PinPoint 3D [5] uses radio frequency lateration to provide an accuracy of three meters and requires a special developed infrastructure. SpotON [6] project developed a 3D location system using RFID tags. SmartFloor [7] uses a sensor grid installed on a floor and the accuracy depends of the sensor spacing distance. These systems require a special infrastructure that needs to be deployed. This is an obvious disadvantage.

Systems that do not require a special developed infrastructure take advantage of an existing and well defined telecommunication network, like cellular or WiFi networks. Cellular networks, like GSM and 3G/UMTS, can provide, with minimal cost, the user current cell identification. The accuracy of this solution depends mostly on the size and density of cells. In urban environments, there is a much better accuracy since more base stations are installed. Signal triangulation is also possible, using the different radio signals from the different base stations [8]. Knowing on which base station the client is connected can also be applied to wireless LANs. This kind of information can be used to locate a wireless client within the access point coverage area.

Implementing a location system using WLAN communication infrastructure has been for some time target of intensive research. Various approaches have been proposed, mostly based on the received signal strength. One of the first systems to be introduced was RADAR [9], developed by Microsoft Research Group. RADAR uses received radio strength to map the user current location and it uses an empirical model (K-Nearest Neighbor) as well as a simple signal propagation model. The accuracy of this system is about four meters for 75% of the time and it uses special developed access points.

R.Battiti et al. [10] describe a system capable of deriving the location using neural networks. In this work it is described the training phase, always present in a RF fingerprinting [11] based system, and the neural network architecture used. Despite the accuracy (about 2.3 meters), the results are only compared to test data and information about software developed, radio strength reading method and system architecture is inexistent. M. D. Rodriguez et al. [12] also describe a location system for hospital services based on a neural network. In this work the SNR is used to calculate the user position and all the processing is done in the wireless client. Their results show an error below 4 meters 90% of the time. Bayesian models are also used to calculate a wireless device position, D. Madigan et al. [13] propose a system with an accuracy of four meters. A. Haeberlen et al. [14] describe a probabilistic approach and their location system provides symbolic information, like the office number inside a building.

There are also location systems that deploy additional hardware, like sniffing devices, on a wireless LAN, with the goal of providing a more accurate solution.

The LEASE system [15] is an example as it uses stationary emitters and dedicated sniffers.

There are several commercial location systems that claim to offer high accuracy positioning solutions. Ekahau [16] uses a probabilistic method described by *P. Kontkanen et al*. [17]. The solution provided by this company is versatile, compatible with different hardware vendors and requires minimum effort for calibration. However, the accuracy is similar to the one that RADAR offered, and the capability of generalization is inexistent.

Siemens has also developed a location system based on neural networks [18] using DECT and WLAN technologies with an accuracy of 5 to 15 meters. Using additional hardware they claim to decrease the error to 15 centimeters, but no information is given on the additional hardware.

Aeroscout [19] uses the TDoA technique to locate special developed wireless tags. This system requires a dedicated infrastructure since several base stations must be installed on the terrain. Systems based on time of flight often provide better accuracy than the ones based on signal strength. This system provides one meter of accuracy, which is more than most applications require. Cisco has recently entered the location system market with a solution named Wireless Location Appliance [20]. The WLA provides positioning using radio strength measurements made by the access points, it only works with Cisco infrastructure and the accuracy is described as “a few meters”.

Analyzing the United States Patent Office database, it is clear that the wireless location paradigm has gained a lot of interest recently. Most of the companies mentioned earlier have one or more patents describing their system, such as

Ekahau [US2005197139] or Aeroscout [US2005207381]. There are also many patents describing various location techniques and systems, most of them from companies like Microsoft [US20060046709], Samsung [WO2006011718], Qualcomm [WO2004016032], etc.

3 Location techniques description

Most common location system techniques are based on time difference. This time difference enables positioning of a mobile device roaming freely on a wireless network. Time difference location systems are based on several methods. In its most simple form, time difference is based on the propagation time of radio signals. The way this time interval is measured or when and where it's measured defines the various methods. The number of actors in the location system can define its accuracy. For example, GPS accuracy is lower when using only two satellites as beacons. Increasing the number of satellites has an immediate effect on the accuracy provided by the GPS.

Other methods of location techniques include angle of arrival (AoA) and radio frequency fingerprinting. AoA uses the antenna aperture angles and azimuths to retrieve the mobile device position. RF fingerprinting requires retrieving received radio signal strengths or other unique radio variables like SNR, and associating these measurements with a position, most of the time described by cartesian coordinates.

Most of the methods can provide enough data to compute the mobile device with enough accuracy depending on the application scope. The way the data is gathered, treated and computed is the main target of a location technique. For example, in RF fingerprinting technique, there are several methods for estimating the mobile device position, from K-nearest neighbor algorithm to artificial neural networks and vector machines.

Nearest Cell

The nearest cell technique is one of the most simple location systems. It can be applied on almost any cell based network, like GSM or UMTS. It can also be used on WLAN, where one can find the nearest access point of a mobile client. This technique does not try to know exactly the client’s location and therefore it does not provide high accuracy as other techniques offer.

In mobile networks, the cell identification is known when a user performs or receives a call.

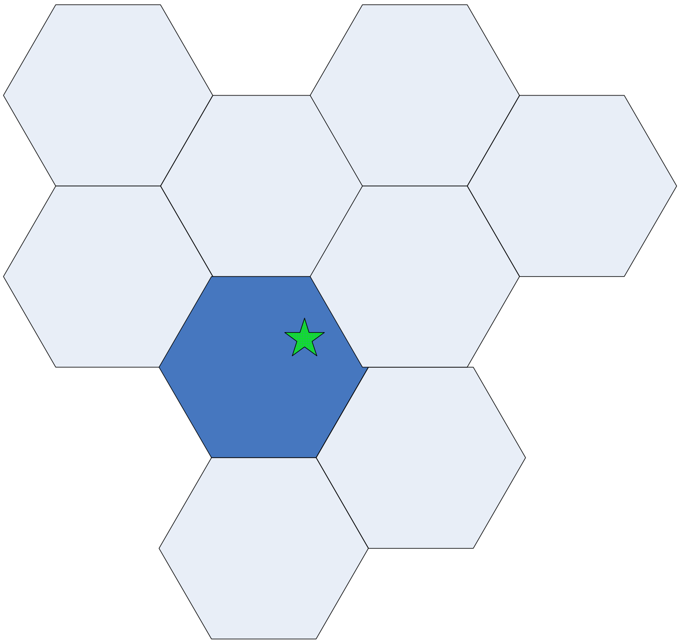


Figure 1- Nearest Cell illustration

This technique merely indicates the cell where the mobile device is associated. Typically, in GSM/UMTS mobile networks, outdoor cell radius can vary from a few hundred meters, like in an urban scenario, to various kilometers in rural areas. In an indoor scenario, where wireless network exists, the access point where the client is associated is indicated. The lack of accuracy of the nearest cell technique is compensated by the very low-cost effectiveness and easiness of implementation. It does not require the use of complex algorithms or calculations. Due to the nature of radio frequency propagation, sometimes the mobile device does not associate to the nearest access point or base station in the case of mobile networks. This disadvantage is minimized by using other type of techniques, like perceiving in which base station the signal is detected with a higher value. Nearest cell can be applied in situation where high accuracy is not a requirement and a simple and low cost solution is enough.

Time of Arrival

The Time of Arrival technique is a simple way of retrieving the wireless device position. The distance between the wireless device and the base station is calculated measuring the time that a radio signal arrives at one point. This point can be the device which is being located or the base station itself depending on the ToA method used. The radio waves propagation speed is a known constant, and knowing the time that a signal arrives, one can retrieve the distance between emitter and receiver, given by the following formula:

$$d_i(x) = v_p t_i \,,$$

where d_i is the distance, v_p is the propagation speed and t_i is time difference between emission and reception of a radio beacon. Considering a complete location system, with several base stations (EB), there are several points for a possible wireless device location. These possibilities are described by a circumference with radius d_i for each base station:

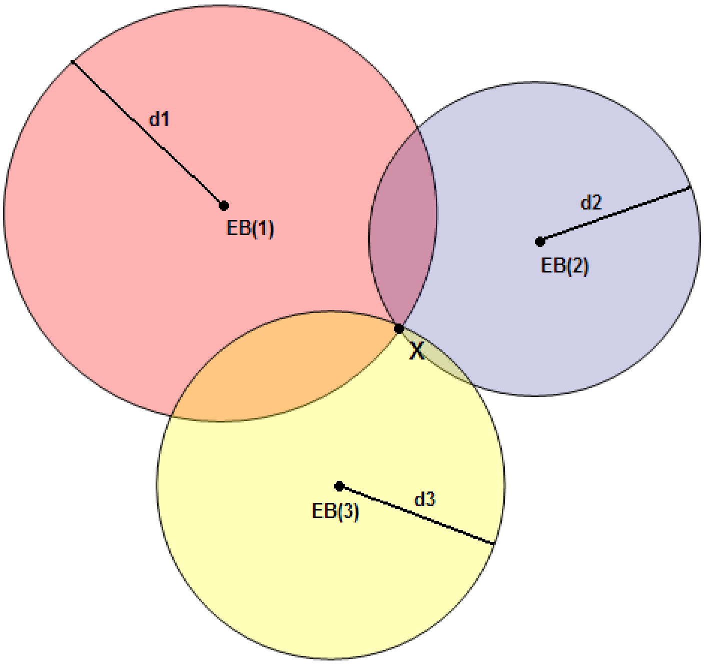


Figure 2 - Wireless client location using Time of Arrival

The intersection between all of the circumferences (X), is the client's location. This method implies the existence of at least three base stations in order to know the exact position of the device that is being located. The main problem associated with this technique is the need of knowing exactly the time that a radio signal is emitted and received. A deviation in the measured time affects the calculated distance, hence increasing the location error. There is also the need to provide a common reference time, so that the exact transmission times are known. In reality most of the times there is not a common intersection point, mainly due to these location errors. In most of the cases, there is the need of performing optimization in order to find the best point, or the most probable intersection point. GPS is a well known example of an application of the ToA technique. The synchronization is assured by using atomic clocks in the GPS satellites.

Time Difference of Arrival

Time Difference of Arrival (TDoA) technique relies on the concept of multilateration in order to locate the wireless client. This location technique measures the time difference of a signal emitted from the client to several base stations. TDoA differs from ToA by using relative time values, where in ToA absolute time values are used.

An emitting signal is sent by the wireless client and received at three or more base stations at slightly different times. These time deviations are due to the different distances that separate the receiver from the several base stations. By using one of the time differences one can calculate the several possible locations of the client. This group of possible locations has a hyperboloid shape and it is described with a hyperbolic function. For example, if a mobile device sends a message, and it is received by station 1 at time T1 and by station 2 at time T2, then the time difference is given by:

$$TDoA_{12} = \left| T_1 - T_2 \right|$$

$TDoA_{12}$ can be used to construct the hyperbola representing all possible locations of the receiver, given by:

$$\left| D_{X1} - D_{X2} \right| = p_{12}(n)$$

Where D_{X1} and D_{X2} represent the distance between base station 1 (EB1) and the mobile client. The constant $p_{12}(n)$ allows the construction of the hyperbola, where each point represents the possible location of the mobile device. Adding a third base station, it is possible to trace a second hyperbola given by;

$$TDoA_{13} = |T_1 - T_3|$$

$$|D_{X1} - D_{X3}| = p_{13}(n)$$

Using several base stations, with a minimum of three, it is possible to acquire the wireless client position by calculation the hyperbolas (one for each time difference) intersection.

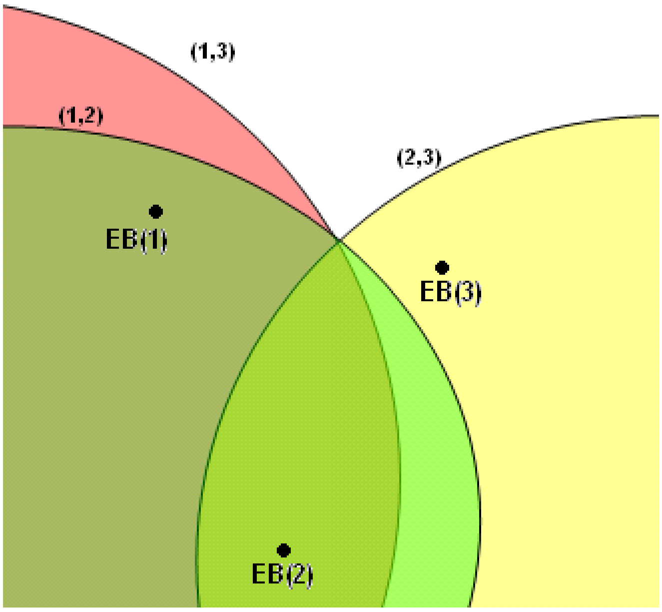


Figure 3 - Wireless client location using Time Difference of Arrival

More base stations can be used in the system improving TDoA accuracy, since more hyperbolas are also used. TDoA is also called Enhanced Observed Time Difference (E-OTD) when the calculations and time measurements are made by the mobile receiver. TDoA has the same requirements in terms of time synchronization as ToA. Recent TDoA systems have developed methods to compensate the errors in the reference time drift, most of the times this is achieved by exchanging a common time reference between base stations. TDoA and ToA require a special infrastructure to operate correctly. For example, in

WLAN access points must be equipped with additional modules in order to retrieve and compute the measured times.

Angle of Arrival

The air interface of a telecommunications radio system is an antenna. These antennas have a known aperture angle and a defined azimuth. When several base stations exist it is possible to pinpoint the wireless device location using the angle of arrival of the radio signal. This angle is acquired from the antenna's azimuth. If the signal is received in different base stations, or antennas, the client's position is obtained from the intersection of the several directions.

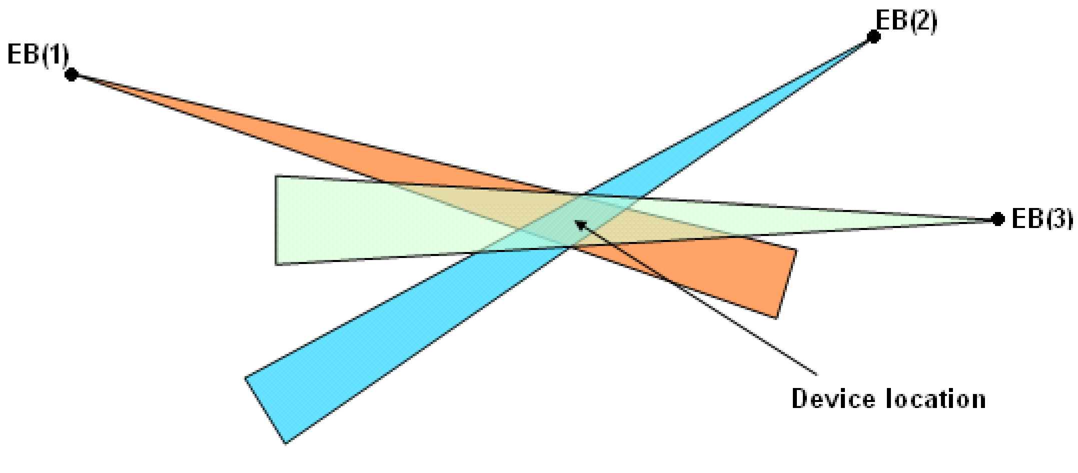


Figure 4 – Wireless client location using Angle of Arrival

In a two dimension plane and if the antennas have enough directivity, only two base stations are required to locate the mobile device. The AoA accuracy is improved by adding more base stations. The existence of a clear line of sight between the mobile device and base stations is one of the requirements of AoA. This method can only be used when the antennas have enough directivity, so it can be possible to define a direction for the signal's source. Also the signal must be received in several base stations, so that a sufficient accurate position can be obtained. The angle of arrival technique can also use more sophisticated antenna arrays. The direction of the radio frequency emitter can be detected with more accuracy by these antenna arrays. Commercial applications of the AoA use this kind of system. In fact, AoA, uses the TDoA between the received signal at different elements of the antenna array. Phase differences at each element of the array provide the necessary data to compute the direction of the signal source.

VHF Omnidirectional Range (VOR) is a well known practical implementation of a AoA system. VOR is used in air industry where aircrafts are equipped with special receivers. Along the surface there are several VOR beacons used by aircrafts to compute their location using the angle of incidence of each beacon signal.

AoA major disadvantage is its need of a clear line of sight. Multipath and reflection effects degrade the system accuracy, and make AoA not suitable for applications deployed in urban or indoor scenarios.

RF Fingerprinting

Radio frequency location fingerprinting does not make use of time measurements, like ToA or TDoA, although merging both the location techniques is possible. RF fingerprinting uses the received signal strength indicator (RSSI) for calculating the device location. This received power is often available in most devices, without the need to use special dedicated hardware, therefore location systems based on RF fingerprinting are often easy to deploy and are cost effective. This technique requires profiling the entire location scenario before the location itself takes place. At each location or point, several measurements are taken and stored. This phase is often denominated calibration. Due to propagation effects, like fading, several measurements are taken to minimize this effect. After profiling the entire area, typically one building with multiple floors, the location system is ready to answer location requests. When a location request is made, new measurements take place and are compared to the ones collected during the calibration phase. The way the data is compared and the location is calculated is the core of the RF fingerprinting technique. Several methods exist like deterministic models (example: Euclidian distance), probabilistic models (example: find the likelihood of a device being at a given point) or other type of models where a simple mathematical approach is not enough (example: neural networks, vector machines, etc.). Each floor or building is defined as being distinct to others in terms of RF signatures. Moreover each location in each building must be unique so that the location pattern can give accurate results. RF fingerprinting technique does not require additional hardware, like TDoA or ToA techniques. It is a technique based on a software solution, when most of the times the RSSI measurements are enough to provide accurate results.

Propagation models

Propagation models technique make use of the received signal strength value to calculate the distance between the receiver and the emitter. When the emitter transmission power is known, one can retrieve the location of the receiver by using several emitters by applying the adequate propagation model. The usage of propagation models to calculate the distance between receiver and emitter in a clear line of sight scenario can be done, in a simple way, using the free space propagation model (Friis formula):

$$PL = 20\log_{10}\left(\frac{4\pi\lambda}{D}\right)$$

The total path loss is given by PL that depends on the frequency used and the distance between the receiver and emitter. The total distance between devices is given by:

$$D = \frac{4\pi\lambda}{10^{\frac{PL}{20}}}(m)$$

Calculating only the path loss is not sufficient, since the receiver is only capable of retrieving the received signal strength. Therefore, for calculating the distance the receiver needs not only to know the RSSI, but also the transmission power. Hence, a more complete way of retrieving the distance is calculating the link budget for the system:

$$P_{RX} = P_{TX} \frac{G_{TX} G_{RX} \lambda^2}{16\pi^2 D^2}$$

Where P_{RX} and P_{TX} are the received and transmission power respectively, G_{TX} and G_{RX} are the gains in transmission and reception and D is the distance. If these variables are known, the distance calculation is given by:

$$D = \sqrt{\frac{P_{TX} G_{TX} G_{RX} \lambda^2}{P_{RX} 16\pi^2}} (m)$$

Using the free space model, the distance to receiver and emitter is known, but for an accurate location, it is necessary to add more base stations, so it can be possible to construct several circumferences of radius D. Their intersection will give the device location. Free space model is not adequate for most of the situations, since propagation is not ideal and other models are more convenient. In fact, the free space model has its own variations, and most of the times its accuracy is questionable. Several propagation models exist depending on the environment where the telecommunication system is deployed. Some are based on measurements taken at several locations, others are pure mathematical calculations. Most of the propagation models can predict the received signal strength at one location with a certain degree of uncertainty. This deviation in values given by propagation models is a barrier for the usage of an accurate location system. In the case of an indoor scenario, this fact is more crucial, since this environment is far from being static and propagation effects like multipath are common. However, using propagation models in a location system is attractive, since they don't require special hardware addition, providing a simple and low cost solution to the location paradigm.

4 Location systems scenarios

Due to the nature of an indoor environment, empirical propagation models are almost impossible to use mainly because the indoor scenario is not static and multi path effects are hard to simulate [21]. Simulation tools can provide a view of the behavior of the signal propagation in an indoor environment. However these tools besides their radio planning usage can not provide an accurate description of the indoor propagation behavior. They require the description of every material present in the building and the use of a complete building layout. These tasks are too complex to be implemented and most of the times this information is not available or difficult to retrieve.

There are different concepts behind location system architecture. The division between different concepts is mainly based on the place where the measurements are done and where the data processing is done.

Concept	Definition
Remote positioning	Measurements are made by base stations
Auto positioning	Measurements made in the wireless device
Indirect remote positioning	Measurements made in the wireless device and sent to the base stations
Indirect auto positioning	Measurements taken by the base stations and sent to the wireless device

Table 1 - Location systems different architectures

The location system developed on this work uses a software approach. The measurements are made by the client device and sent back to a server using the wireless access points. The system uses *indirect remote positioning*. This solution takes the processing load to a central server and the client just needs to collect some measurements and send them to the server.

5 System description

Description

The location system developed is divided into two different sub-systems. The first sub-system named macro-location has the objective of retrieving the information about the user current area or room in a building. This kind of information can be gathered by methods such as Cell ID (in case of pico-cells) from GSM networks or associated Access Point from WLANs. Macro-location’s main goal is filtering out unnecessary areas and pointing to a defined zone or room. The other subsystem, micro-location, provides a more accurate location, within a few meters and uses radio frequency fingerprinting for pinpointing users location. RF fingerprinting was chosen due to its flexibility and robustness in indoor environments. Other techniques like ToA or TDoA must use dedicated hardware in order to perform the required time measurements. Most of the time, this hardware must be powerful enough to perform time measurements accurately. Factors like sampling times, deviation errors and environment influence play a crucial role. The requirements of this location system state that, it must be easily deployed on a real scenario. Using additional hardware raises the system complexity, so a software solution is the answer. Common wireless devices have access to signal strength measurements, like the case of Bluetooth or WiFi clients. These signal strength values can be used by special developed software in order to locate the wireless client. Propagation models could also be used to try to locate the wireless devices, but most of the time, they are not accurate in indoor environments.

Requirements

An indoor location system must be simple to implement in real life. In other words it must be low cost, easy to interface with and provide accurate results. Most common indoor radio source devices that exist in indoor environments include Bluetooth handsets and wireless access points. Due to the availability of these devices in most buildings, Bluetooth and WiFi are clear candidates for being used in an indoor location system. These technologies are sufficiently mature, widely available and a low cost solution. In order to not increase the cost of a final solution no additional hardware should be used. The location system should be based on a software-only solution. This way the integration with different kind of vendors and equipment must be easily enough so that the system can be scalable.

Requirement	Description
Flexibility	Able to be installed in all WiFi infrastructures available on the market
Low cost	Make use of a software based solution avoiding the need of specialized hardware
Scalable	Independent of the number of clients used and of the size of the location scenario target
Generalization	Provides positioning information in places where an insufficient calibration was performed
Accuracy	Capable of providing accurate locations. Average Error < 2 meters

Table 2 - System requirements

The system should also be simple enough to be fast and with high usability. It can also be adapted to different kind of needs and requests. Hence the accuracy can be adapted to the needs of the user and if the location speed is a requirement, it should also be changeable.

6 Macro-location

In a typical office wireless LAN there are several access points distributed in a defined way in order to offer good signal reception and sufficient bandwidth. Knowing which client is associated with which access point plays a critical role in the *macro-location* scenario. In fact, the main objective of the macro-location is to retrieve the location of the mobile client using the access point coverage. Macro-location is an application of the *Nearest Cell* technique described earlier. Despite the existence of several ways of knowing in which access point the client is associated some disadvantages can be found in each one of them. There is the need for a general solution that can be easily deployed to the existing variety of vendors and manufacturers.

SNMP – Simple Network Management Protocol

Access points act as a transparent bridge forwarding packets from/to wireless clients. Using SNMP (Simple Network Management Protocol) it is possible to manage all the SNMP-enabled devices on a network in an efficient way. The efficiency of the network management depends on the information available on dedicated information tables called MIB (Management Information Base) that are located on network devices. WiFi devices that support SNMP use a MIB called ieeedo11-Mib. This MIB has information concerning the access point's configuration. Using SNMP and accessing the access point MIB it is possible to acquire client information [22].

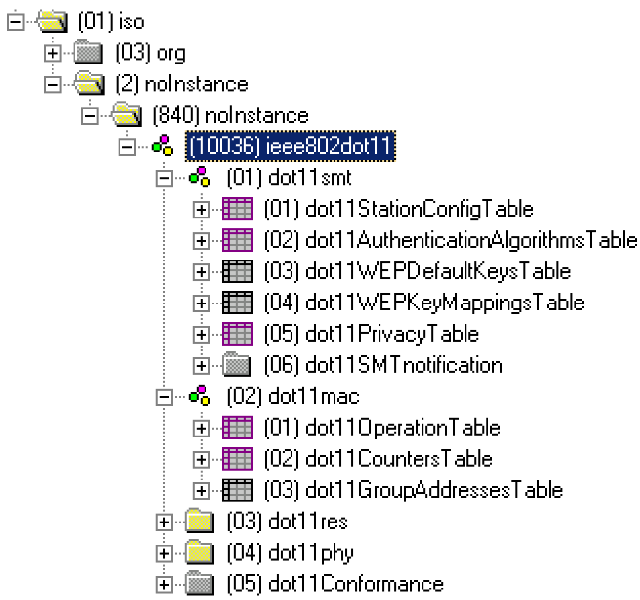


Figure 5 - ieeedot11-MIB available in some access points on the market

Unfortunately, not all vendors have SNMP support and this solution requires pooling each access point for a list of associated clients or to configure each access point to generate SNMP alarms.

Telnet

Other solutions proposed included telnet access to retrieve information about the clients that are associated on an access point. This method implies accessing the access point via Telnet and sending commands to acquire the necessary information. The number and type of commands that are required depend on the model and vendor of the access point. Clearly, there are some limitations in the usage of this method. Telnet access requires knowing the authentication credentials like username and password. It is also required to perform the necessary commands each time a user demands a location request to be able to provide real time information.

```
ap#show dot11 associations
show dot11 associations
802.11 Client Stations on Dot11Radio0:
SSID [Local-wifi] :
MAC Address      IP address      Device   Name    Parent   State
0002.2d64.40f7 193.136.93.100      -        -      self    Assoc
Others:  (not related to any ssid)
```

Figure 6 - Example of Telnet commands retrieving client association data sent to a Cisco 1100 Access Point

These commands are sent to all access points where the macro-location is deployed. Considering a network with several access points, the location process can be slow. Also telnet commands are dependent on vendor, model and firmware version. This requires making a huge effort to gather all types of commands and their variations.

Dedicated software

Another solution proposed to solve the collection of clients associated with APs, was to install dedicated software on each wireless client. This software would be resident on each device and would listen to a specified network port waiting for a macro location request. These location requests would be sent by a dedicated server. After receiving a request, the special developed software solution would send a message informing on which access point the wireless client was associated.

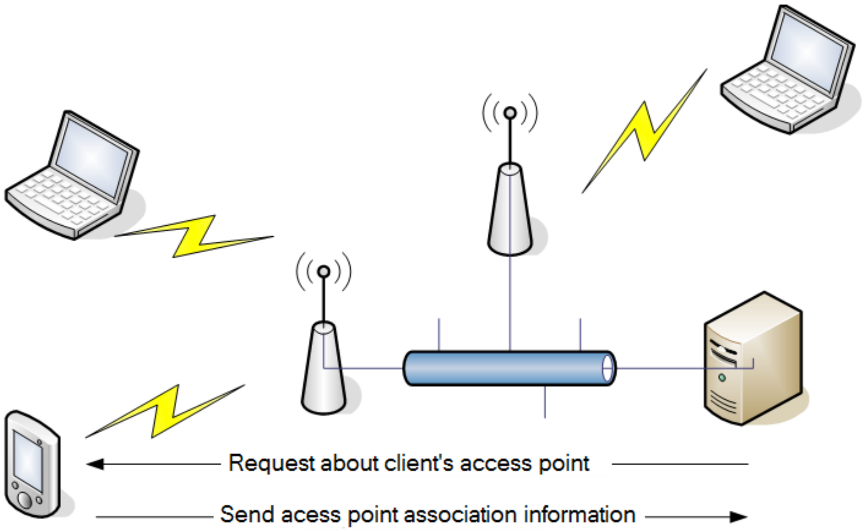


Figure 7 - Dedicated software macro-location architecture

This type of solution is unpractical since it is required to have software installed in each wireless client. Considering the goals of the macro location system and the other type of solution available, dedicated software is far from perfect.

WMI – Windows Management Instrumentation

Windows Management Instrumentation (WMI) is a management technology allowing scripts to monitor and control managed resources throughout the network. Resources include hard drives, file systems, operating system settings, processes, services, shares, registry settings, networking components, event logs, users, and groups. WMI is built into clients with Windows 2000 or above, and can be installed on any other 32-bit Windows client.

Remote procedure calls (RPC) is a technique used in client-server applications. RPC extends the notion of conventional local procedure calling where the called procedure is not required to exist in the same address space as the calling procedure. These two processes can be on the same system or in different systems using a network to communicate. The transport mechanisms of RPC make it easier for the programmer to use distributed applications in which the network itself is transparent.

Using WMI and Remote Procedure Calls it is possible to acquire all types of information remotely. This includes details about wireless information, in other words, the access point that a client is associated.

WMI is a Microsoft Windows based solution, therefore all the other operative systems devices are excluded by using this type of solution.

Considering a location scenario, a server would send a request to a client using a remote procedure call. The information returned by the wireless client can contain several data like the access point mac address, SSID, received signal strength, etc.

```
Set fso = CreateObject("Scripting.FileSystemObject")

strComputer = "193.136.92.156" ' ip or host name
userName = “guest"    ' user
passWord = “guest"    ' password

(…)

set colInstances = objSwbemServices.ExecQuery _
  ("SELECT * FROM MSNdis_80211_ReceivedSignalStrength WHERE Active = 
True AND InstanceName ='" & wifiAdapter & "'")
for each objInstance in colInstances
        sigraw = objInstance.Ndis80211ReceivedSignalStrength
        signal = sigraw & "dB"
Next

(…)

Wscript.echo "Client associated to:"& vbcrlf & "mac: " & mac & "  SSID: " & ssid & "
RSSI: " & signal
```

Figure 8 - Example of a WMI script access wireless information from a remote computer

This type of solution has several limitations including the fact that it is mandatory that the server has permission to access a remote machine. Network architecture and firewall existence can compromise the use of remote procedure calls, therefore making impossible the connectivity between client and server.

Authentication

The authentication process occurs in a wireless association when the access point demands that the client presents security credentials. In large wireless networks, this authentication is made using a Radius server that manages all the security of the wireless network. This kind of data exchange could be captured and analyzed to return the access point where the association was performed. Another approach is to use the Radius server itself to retrieve this information.

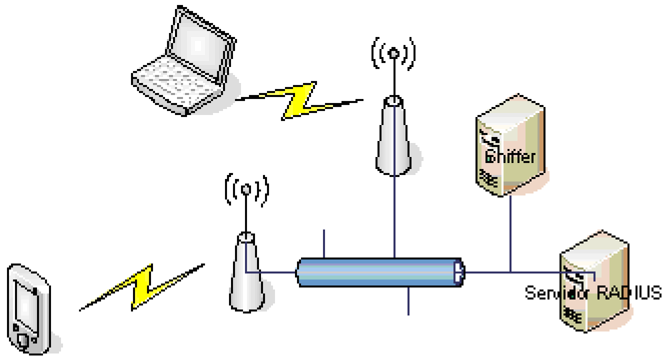


Figure 9 - Authentication architecture

This method requires major changes in the wireless network behavior. Performing changes in a Radius server or access security data is not simple to achieve. There is also the requirement to use authentication in the association process. Some networks do not use Radius authentication and make use of simpler methods like WEP or WPA.

Syslog

The final solution is based on the remote syslog [23] capability of most access points in the market. Remote syslog allows a device to send important messages to a central server that gathers all the data.

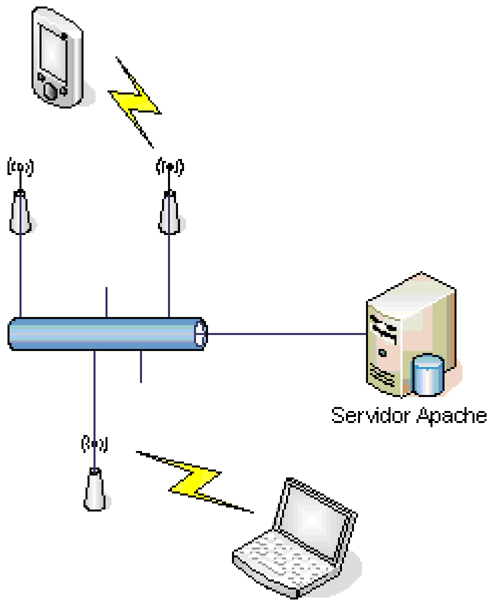
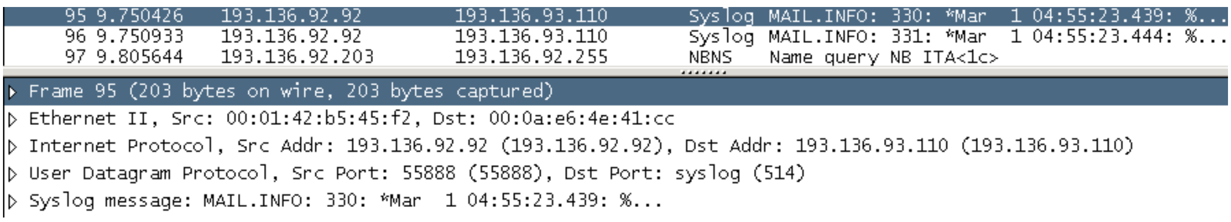


Figure 10 - Macro - location system architecture

With this method when a client associates/disassociates, the access point informs the server of this event by sending a syslog packet. The server has specific developed software in order to treat these messages and collect them into a database. Then it is possible to search for the location of a client or to see which clients are in an area.

Sylog standard packet description is described by RFC 3164. The maximum payload size of a packet is 1024 bytes. However, in some traffic sniffing performed it was noticed that the average payload size of the syslog packets sent by access points when a client performs an association, was between 160 and 200 bytes.



```
95 9.750426   193.136.92.92    193.136.93.110   Syslog MAIL.INFO: 330: *Mar  1 04:55:23.439: %...
96 9.750933   193.136.92.92    193.136.93.110   Syslog MAIL.INFO: 331: *Mar  1 04:55:23.444: %...
97 9.805644   193.136.92.203   193.136.92.255   NBNS   Name query NB ITA<1c>

▷ Frame 95 (203 bytes on wire, 203 bytes captured)
▷ Ethernet II, Src: 00:01:42:b5:45:f2, Dst: 00:0a:e6:4e:41:cc
▷ Internet Protocol, Src Addr: 193.136.92.92 (193.136.92.92), Dst Addr: 193.136.93.110 (193.136.93.110)
▷ User Datagram Protocol, Src Port: 55888 (55888), Dst Port: syslog (514)
▷ Syslog message: MAIL.INFO: 330: *Mar  1 04:55:23.439: %...
```

Figure 11 - Syslog Packet description

A large wireless network can have several access points. It is important to study the effect of additional traffic and its impact on the network. The values presented above are based on the worst case scenario. It is presented the number of syslog messages for each minute in each access point versus the necessary bandwidth

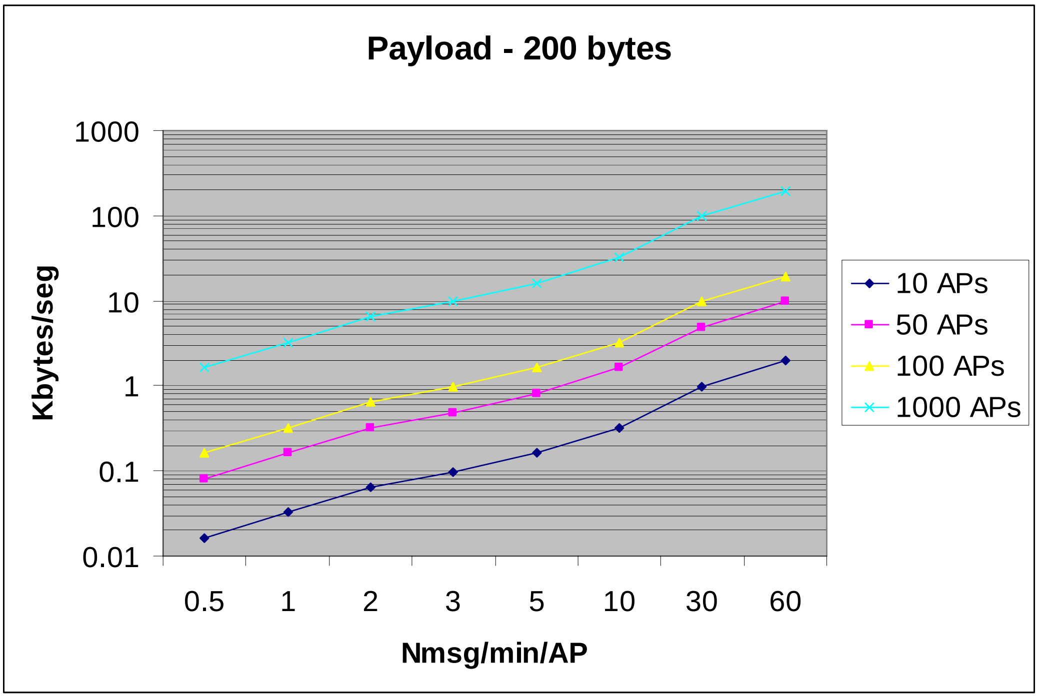


Figure 12 - Syslog messages load - Payload size 200 bytes

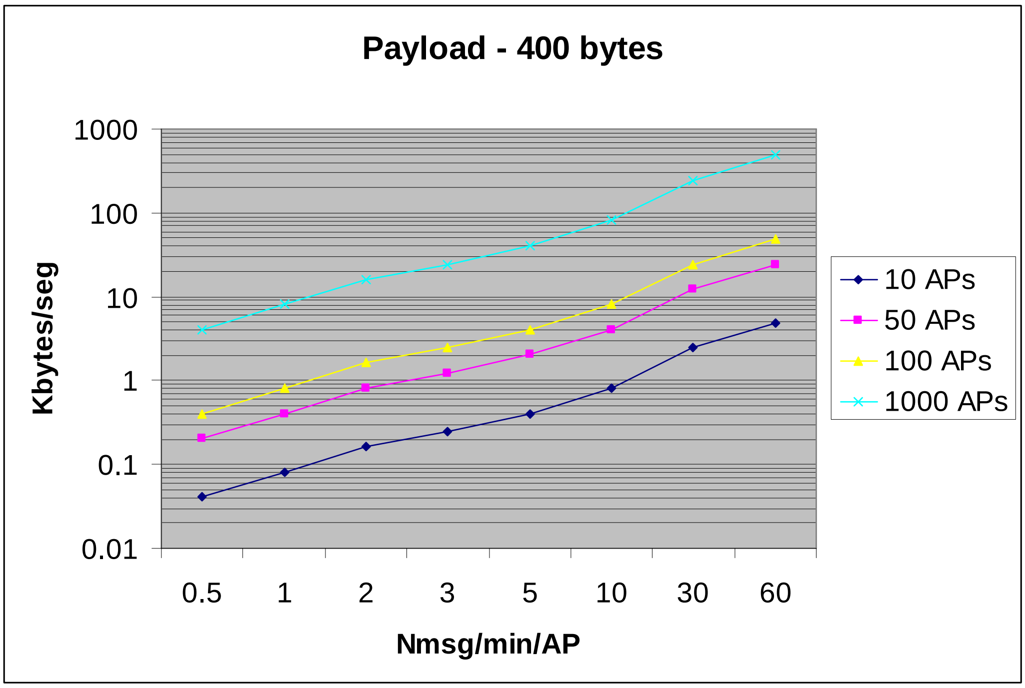


Figure 13 - Syslog messages load - Payload size 400 bytes

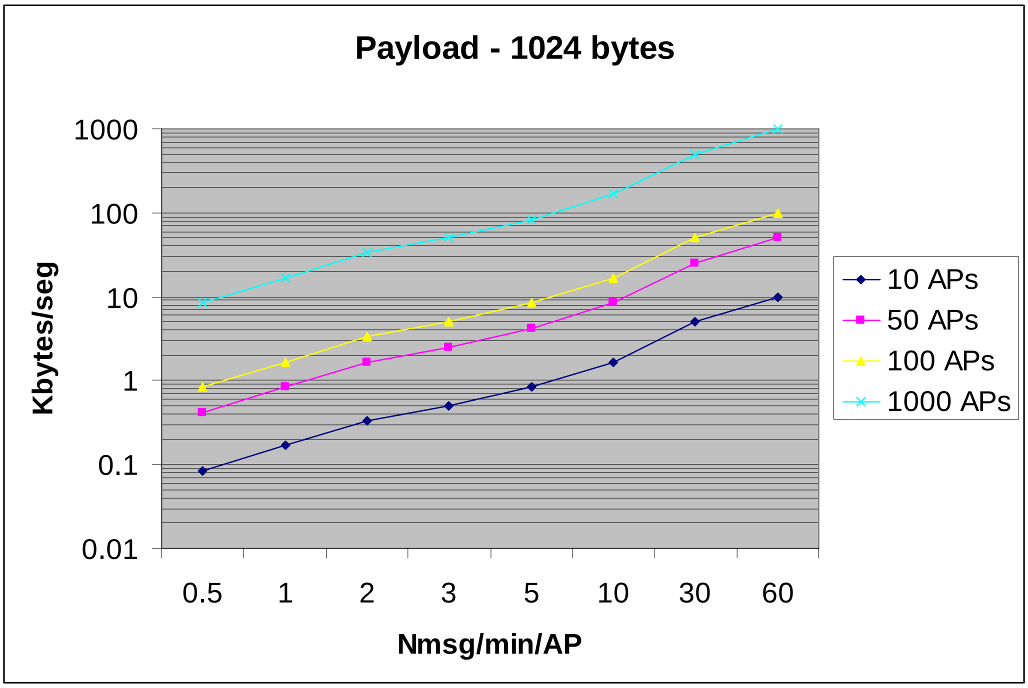


Figure 14 - Syslog messages load - Payload size 1024 bytes

A message is sent each time a wireless client performs an association or disassociation. The number of this events per minute basis on a typical network should be high. Hence, the values presented on these graphics are excessive comparing to a normal reality. In a scenario with 1000 access points and two associations per minute in each access point, there would be a requirement of 30 Kbyte/s of bandwidth. Considering a typical wireless network backbone (100/1000 Mbits), this amount of traffic does not make a huge impact.

This system was tested in the *Instituto de Telecomunicações* building using three Cisco access points. These APs were configured to send all the logs to the central server. Various location scenarios were tested, including fast moving clients, turning off one or more APs and also turning off the client. In all cases the location system provided the location information correctly. Increasing the number of access points has little effect in the system. The network performance is unaffected since the log packets are small and not in sufficient number to cause a major impact on typical office LAN.

7 Micro-location

Most common location systems require a special infrastructure. Others also require the monitored device to be modified by adding special hardware. Clearly, a software based approach has the advantage of being easily deployed since it does not require modifications on the wireless device or the existing infrastructure. Also, it has the ability to be easily adapted between equipment manufacturers and platforms. The location system must also provide accurate position information. The measurements can be done in the infrastructure or by the client. Since there is a standard way of performing signal measurements on the client side, it is easier to develop a software solution that can be deployed almost at any type of wireless device. An infrastructure solution has the disadvantage of being dependent on manufacturer implementations, since there is not a standard way to retrieve radio strength values from the various types of access points.

Location Algorithms

The location system developed makes use of the radio frequency technique to pinpoint the exact location of a mobile device. As described earlier RF fingerprinting has two distinct phases:

- Calibration phase
- Operational phase

This section describes the steps performed during both phases. For constructing a radio frequency pattern of the building where the location system was applied, two different types of software were developed. The first is more accurate, since the user is required to insert the exact location in spatial coordinates. The second uses another interface approach, displaying an area map, where the user can pinpoint his location by clicking on it.

During the operation phase, several algorithms were tested in order to discover which one was more accurate and reliable.

System Calibration

Radio frequency fingerprinting makes use of signal strength measurements to describe the characteristics of propagation inside a building. This method relies on taking several readings at different points. Each point may be defined as Cartesian coordinates or by symbolic information. The signal strength vector associated with each point can change dimension according to the number of signal sources which are being measured. In order to fully characterize an environment full set of calibrations must be performed. The granularity of points chosen will influence the location system performance. Higher granularity often means better accuracy.

If there are *m* points, then:

$$p_i = (x_i, y_i), i = 1..m \quad (1)$$

The point position is described by its coordinates (*x,y*). It is also possible to add a third dimension, namely when the scenario involves multiple floors inside a building. In this case, a coordinate *z* is added that corresponds to the floor number. At each point, the received signal strength from each base station is measured. Considering that *n* base stations exist, then the signal strength vector (*R*) is defined as:

$$R_i = (a_{i1}, a_{i2}, ..a_{in}) \quad (2)$$

Where a_{ij} corresponds to the measurement collected from point *i* from base station *j*. The complete RF fingerprinting matrix (*M*) is given by combining spatial information with the RSS vector:

$$M = \{p_i, R_i\}, i = 1..m \quad (3)$$

Combining with (1):

$$M = \{(x_i, y_i, R_i)\}, i = 1..m \quad (4)$$

Collecting signal strength measurements in a large area is time consuming, however it is only required once if the site layout doesn't change. In order to collect measurements, in an accurate way, several techniques can be used. Profiling the area with a grid and following it while taking signal strength readings is an accurate method but requires marking points on the floor or the use of other similar techniques.

Figure 15- Grid calibration method - outdoor scenario tryout

Calibrating the location system using a grid can be an enormous task when the location scenario, like a building, has a considerable size. The first try out was to use a grid deployed on an outdoor scenario with line of sight with three dedicated access points.

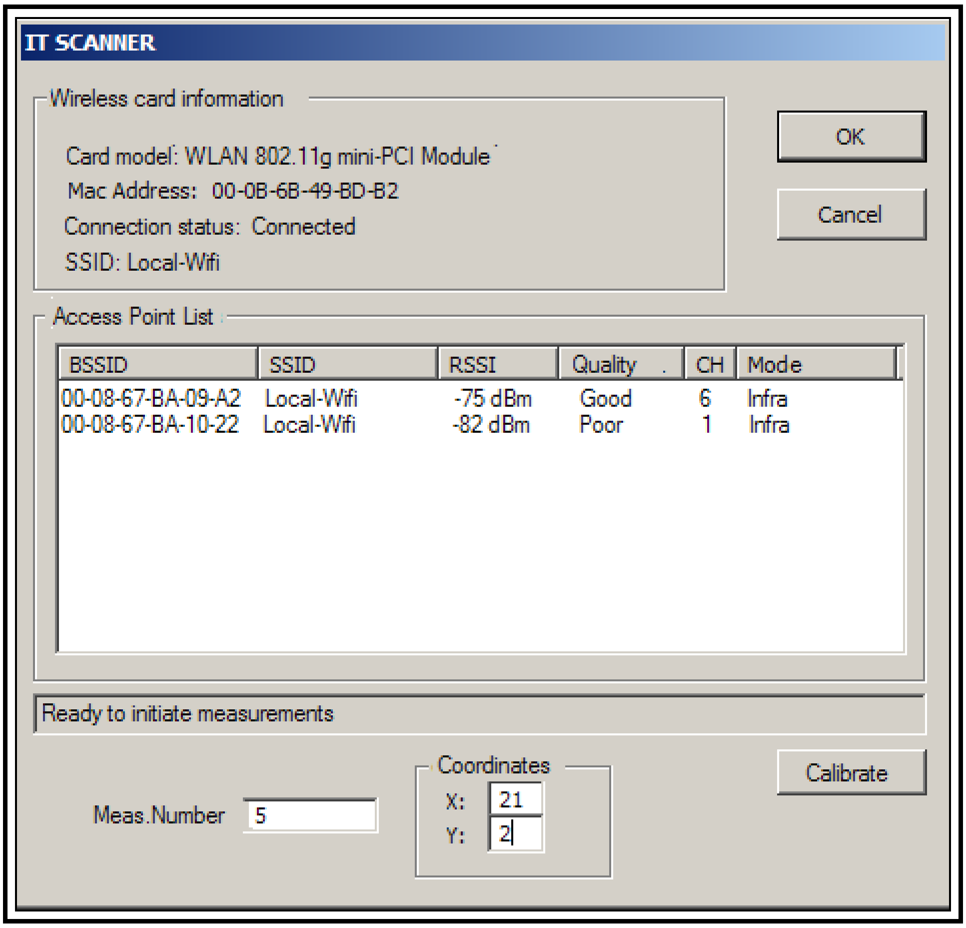


Figure 16 - Calibration Software - Grid based

Another efficient method is importing the floor plans to a special developed software (or web application), where the user can click on the desired measurement location. This method can be less accurate, but it is faster and makes the process of collecting measurements easier.

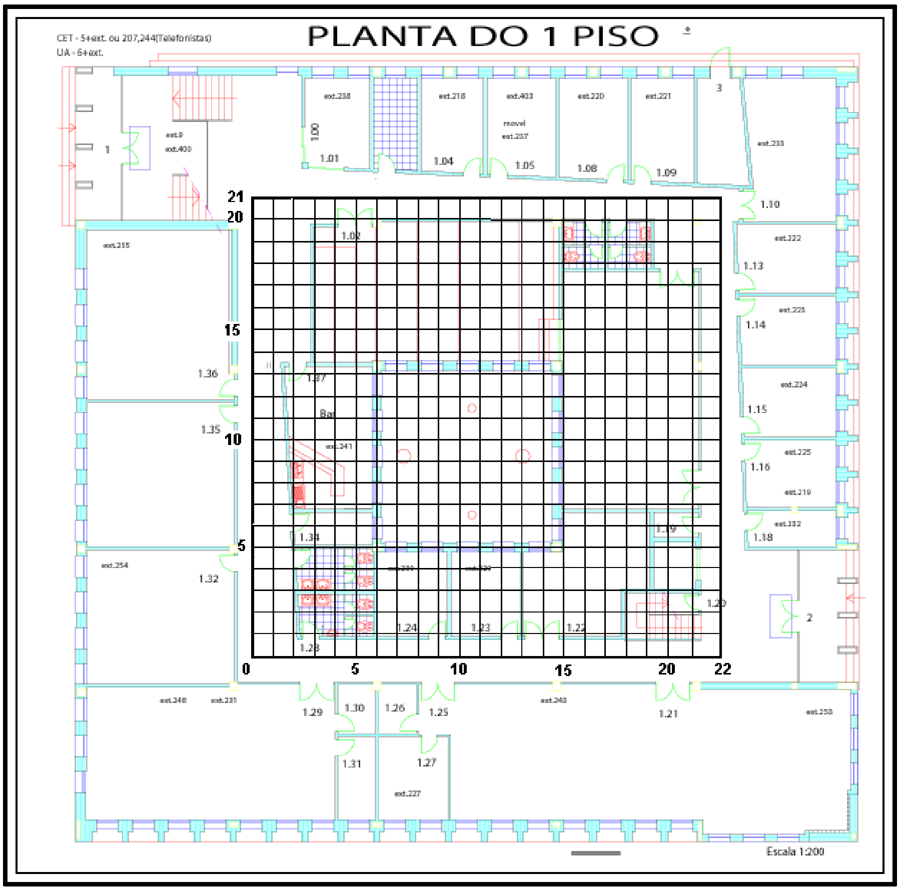


Figure 17 - Example of a grid applied to the floor plan of *Instituto of Telecomunicações* first floor

Signal strength values vary according to the device orientation. This effect is more notorious when using devices equipped with non-omni directional antennas (example: Laptops, PDA, etc). A robust location system must take in account these details, since they produce a profound effect on the system performance. While taking measurements, a full 360º rotation can be made to ensure that all the directions are covered. This way, the effect of the wireless device orientation is minimized. Considering that several measurements are made in the same point, each a_{ij} defines the signal strength behavior. The value of a_{ij} can be calculated using the average signal strength of each base station.

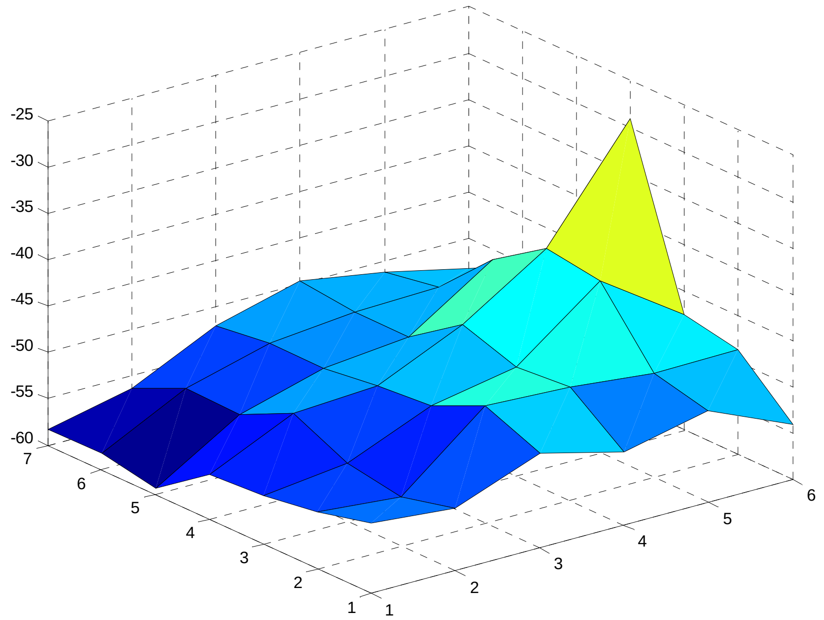


Figure 18 - Signal strength distribution from one access point

Algorithms

The location technique used by this system, radio frequency fingerprinting, requires a previous calibration. When the locating mechanism takes action it is necessary to retrieve signal strength values and compare them with the calibration data. This comparison can be performed by various ways, like calculating the Euclidean distance between signal strength values used in calibration and the ones retrieved from the wireless client or using propagation models. These methods were the first approaches for being the algorithms of choice of this location system. Despite being fast algorithms they lacked the capability of generalization. Artificial neural networks were the final choice for location algorithm. They provided the flexibility, speed and generalization abilities required by the location system.

Propagation models

The first step for developing the location algorithm was the usage of propagation models. The first approach was to collect measurements by a mobile device in an outdoor scenario with clear line of sight to the base stations, in this case wireless access points. After collecting the measurements in several points on a grid, several simulations were performed in order to study the accuracy of several propagation models. A free space model variation was applied to the measurements, since both the client and access points' location was known.

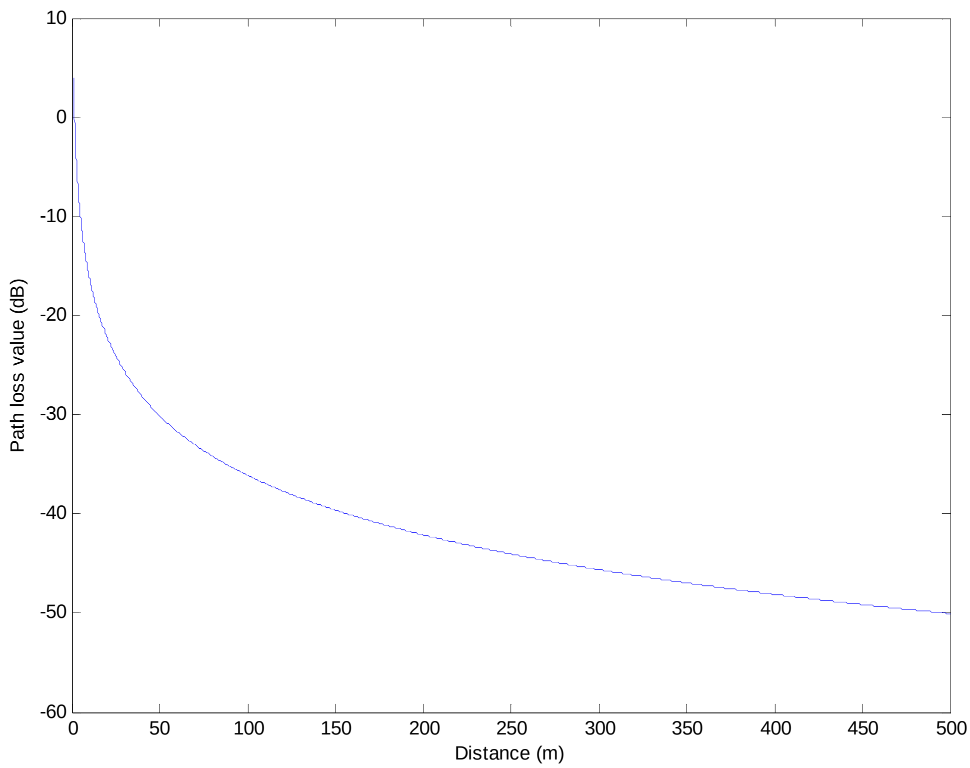


Figure 19 – Theoretical free space loss variation over distance

One of the main problems of using propagation models is fading and the mobile device orientation. In the experiments that took place it was clear that the mobile orientation had a major influence on the collected RSSI values. Applying theoretical propagation models to the collected data did not give accurate results. Despite the efforts to adapt various existing propagation models, the average error was higher than the location system initial requirements. Moreover, by using propagation models, the location system had not the capability to adapt to environment changes. Take for example a variation of four to five dBs, (typically a fade margin). In this case, the distance between emitter and receiver could change by several meters, decreasing the system's accuracy dramatically. The empirical propagation models tryout lead to other approaches, like the nearest neighbor.

Nearest Neighbor

After the acquisition of sufficient signal strength data, it is necessary to develop a system capable of locating a wireless device. This system must use the collected data and try to construct a predictive model capable of performing real time tracking of a wireless device. This tracking has two distinct phases:

1. Request signal strength data from the wireless device
2. Use requested data to calculate the device's position

Calculating the position of a wireless client can be a serious challenge. There is the simple possibility of making a comparison between the signal strength data collected in the calibration phase. This can be done using a simple algorithm like the nearest neighbor rule. Take for example a signal strength vector request from the wireless client such as to calculate its position:

$$m = (b_1, b_2, ..., b_n) \quad (5)$$

Then, the device's location coordinates, *x and y,* will be the ones that give a minimum:

$$\min\left[d(a_{i1}, b_1) + d(a_{i2}, b_2) + ...d(a_{in}, b_{in})\right] \quad (6)$$

After finding *i*, the *x* and *y* coordinates are given by (1). Despite giving some interesting results, this method has some disadvantages. There is no capability of predicting the client's position outside of the calibration points. In other words, when a client is between some of the grid points, the final location given by this method is not error free.

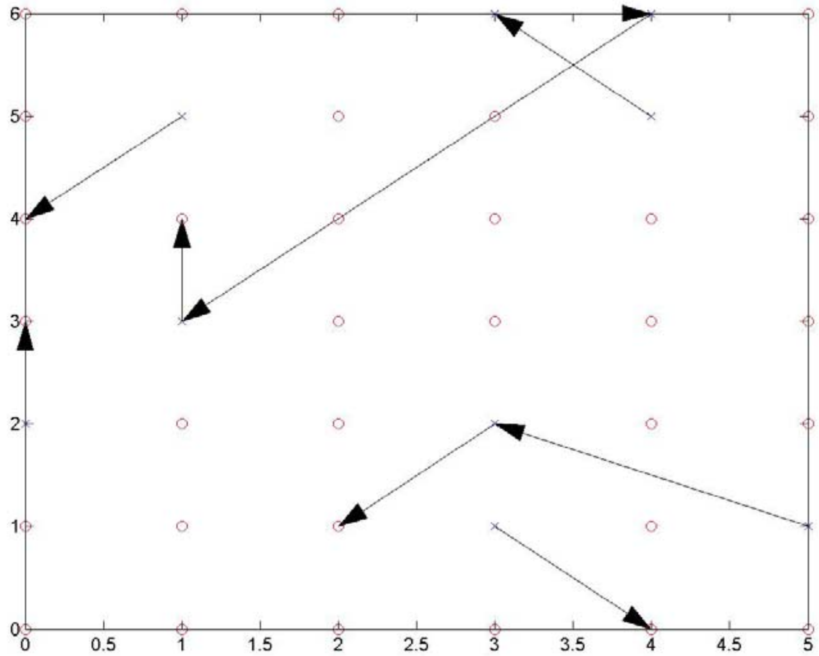


Figure 20 - Grid used in testing the nearest neighbor algorithm (arrows correspond to location errors)

Since indoor environment is not static, RSSI variations exist at the same point but at different times. A location system that uses the nearest neighbor algorithm is critical affected by these variations. Clearly, there is the need for a more adaptive method.

Artificial Neural Networks

The lack of generalization by the nearest neighbor algorithm is a disadvantage. In order to try to achieve better results, some tests were performed using neural networks. Artificial neural networks (ANN) have the ability to learn with a defined data set [24]. In this case the behavior of the signal strength of each access point is learnt by the ANN. The data collected during calibration is used to train the ANN. After this training, the ANN can be used to calculate the wireless client position. The type of neural network used was the multi layer perceptron (MLP). MLP have the ability to classify data patterns and also to approximate mathematical expressions efficiently. As the name suggests, MLP are composed of different layers: an input layer, one or more hidden layers and an output layer. Each layer is composed of several neurons that have a bias and an activation function associated. The inputs for the ANN are the signal strength measurements taken in the calibration phase. The output is the position of the device, described by *x* and *y* coordinates.

The number of inputs may vary according to the number of base stations used, a value between three and five is considered adequate. The input and hidden layer activation functions are based on the hyperbolic tangent sigmoid function:

$$f = \frac{e^{n} - e^{-n}}{e^{n} + e^{-n}} \qquad (7)$$

The output layer activation function is linear, since the outputs are spatial coordinates. The number of outputs is typically two, since a two-dimensional plane was used to describe the possible locations of a device in an office floor. Nevertheless, it is possible to adjust the neural network to accommodate an additional output, this way a third dimension is available (example: number of the floor).

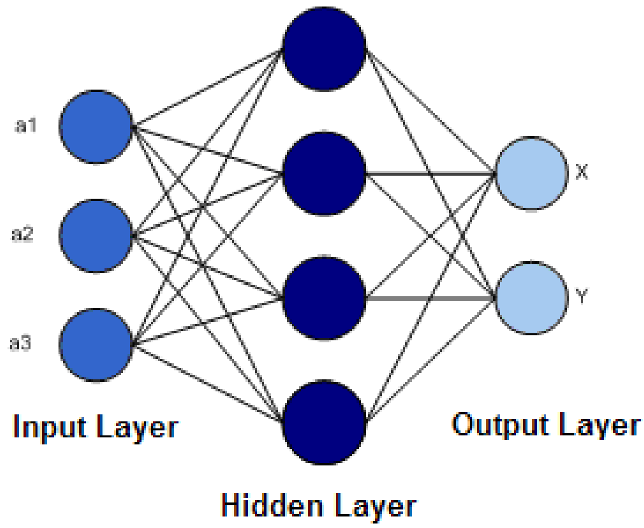


Figure 21 - Artificial Neural Network general architecture

The number of neurons in the hidden layer is variable and can be adjusted to the size and behavior of the training data. Typically, a large set of training data with different input/output values requires an increase in the number of neurons. The training process of a neural network must be adequate so that problems like over-fitting should not arise. Over-fitting occurs when too much training is applied to the ANN. This means that the ANN will be fitted exactly to the training data, therefore losing all the generalization capabilities. On the other hand, a poor training makes the ANN not to learn adequately

The number of inputs, or base stations used, also has an impact on the training performance. An increase in the number of base stations has the advantage of decreasing the expected average error, but there is also the need to increase the number of neurons, giving additional complexity to the location system.

Practical Implementation

Towards a final solution

The first step made to use neural networks was to use Matlab © Neural Networks ToolBox to provide an insight in whether ANNs were or not a good solution to the location problem. Radio signal strength data was collected from a corridor in Instituto de Telecomunições. This data was composed from RSS from three access points that existed in the IT's main building. In this first step, only one dimension was used, this means that the output's size from the ANN is only one. At each point of the corridor ten measurements were made in different directions.

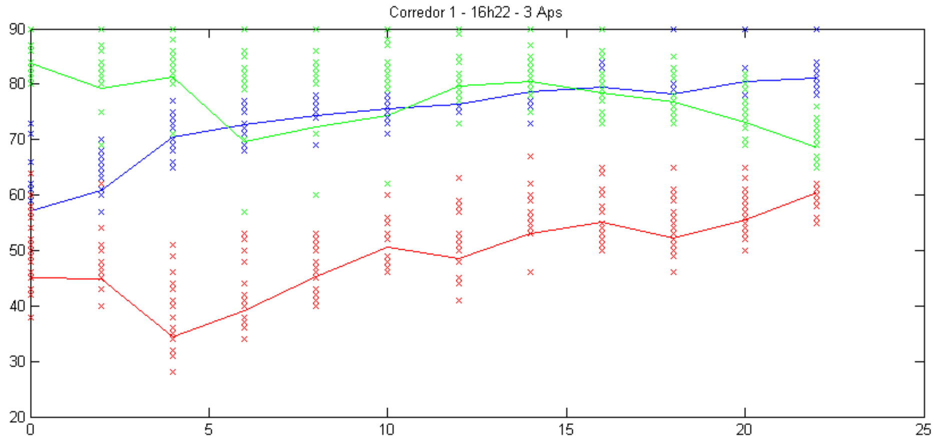


Figure 22 - RSS variation along a corridor for three access points

At this point there is more than one solution on how to treat the RSS data in order to better suit the ANN. As expected, the RSS value is not constant at the same point. This is mainly due to the change of the device direction and also to the nature of the indoor radio propagation channel. As there are ten measurements for each point with different values, it is possible to take random readings and use them to train the ANN. Another approach is to calculate the average RSS value for each point. The use of random points has the clear disadvantage of not providing an accurate description of the signal behavior. Also the ANN training is more difficult since for the same output there are different input values. The solution was to simply take the signal strength average value for each point. This way the signal strength vector used, acquired during calibration, becomes:

$$R_i = (\overline{a}_{i1}, \overline{a}_{i2}, .., \overline{a}_{in}) \quad (8)$$

Other methods may be applied with similar final results, such as a moving average.

ANN creation

The construction of a final neural network system involved the analysis of several software packages, since Matlab NN ToolBox was not flexible enough for the goals of the location system. Figures of merit analyzed included flexibility, speed of training and execution, capability for running in several programming languages and power to scalability. The final choice was Fast Artificial Neural Network Library (FANN) [25]. This library includes several popular training methods, various hidden functions and the capability for being used in several programming languages and platforms. FANN also allows an easy integration with any type of software that is being developed.

Several parameters exist in an ANN. These parameters define the behavior and performance of the ANN. The size of the area and the granularity of points where the location system is deployed have a profound influence on the ANN training duration and on its structure complexity. Several structures were tested but the ones that showed better results were N -> 8 -> 2; N-> 12->2 and N->16->2. The connection rate between neurons was chosen to be one, since changing to a lower value and keeping the same number of hidden layer neurons degrades the performance of the ANN.

ANN hidden function

Performing a decent training plays a crucial role in the future performance of the ANN. There are different types of training methods that are described in [24]. The criteria to choose the best training, was the one that showed better results. In order to study the best training method and hidden function, several RSSI values were collected. Two data sets were gathered, one for training and another for testing. Also several different variations of the sigmoid hidden function were tested:

Training method	Function	Error (m)
RPROP	SIGMOID	3,454
RPROP	SIGMOID_STEPWISE	2,040
RPROP	SIGMOID_SYMMETRIC	2,350
RPROP	SIGMOID_SYMMETRIC_STEPWISE	2,884
QUICKPROP	SIGMOID	2,122
QUICKPROP	SIGMOID_STEPWISE	2,206
QUICKPROP	SIGMOID_SYMMETRIC	1,835
QUICKPROP	SIGMOID_SYMMETRIC_STEPWISE	1,521
BATCH	SIGMOID	2,313
BATCH	SIGMOID_STEPWISE	2,272
BATCH	SIGMOID_SYMMETRIC	1,745
BATCH	SIGMOID_SYMMETRIC_STEPWISE	1,843
INCREMENTAL	SIGMOID	2,189
INCREMENTAL	SIGMOID_STEPWISE	3,465
INCREMENTAL	SIGMOID_SYMMETRIC	4,938
INCREMENTAL	SIGMOID_SYMMETRIC_STEPWISE	4,979

Table 3- ANN training method and hidden function comparison

These tests showed that the best training method was Quick Propagation with a final training error of 1,521 meters. Quick propagation is variation of the backpropagation algorithm [26]. Backpropagation is based on propagating the output error back to the network when an input is applied. The weights and biases of the ANN are adjusted to minimize the mean square error. The training functions for these tests were all based on the sigmoid function. The symmetric variation has a negative bias to the outputs. The stepwise sigmoid, as the name suggests, implements a fixed point function in a non continuous way. This study provided a sufficient insight on the behavior of the several training methods when applied to the location system.

ANN structure and parameters

Another significant parameter in artificial neural networks is the learning rate. It affects the learning capability of the ANN, and a suitable value is required to perform an adequate training. The correct setting of the learning rate is often dependent on the size and type of input data and is typically chosen through experimental testing. Its value can also be adapted during the training phase, therefore becoming time dependent. The value for the learning rate chosen was 0.01. This low value is related to the nature of the input values that were between -1 and 1.

The location system is dependent on the number of base stations detected, in this case the number of access points. While collecting measurements in the area, the number of base stations is not important. But before training the neural network, a decision must be made concerning the number of access points used. In a typical medium-size office building covered by a wireless network, there should be a significant number of access points. In order to adapt the location system to reality, the system uses between three or four access points, depending on the size and nature of the location area. However, some experiments were made to test the effect on raising the number of access points used.

The number of base stations used is also the number of inputs of the neural network. Using a large number of inputs raises the complexity of the NN. On the other side using a low number does not provide sufficient data, so the NN might not converge to a final and acceptable solution. The choice of which access points to use in a real scenario must be based on prior knowledge of the reliability of its characteristics like location and signal strength. Roaming access points or other untrustworthy sources of wireless signals must be discarded since they are not reliable.

The radio strength values gathered from wireless access points are measured by the client in dBm units. These values can go from 0 dBm to -90 dBm based on experiments made in a typical scenario:

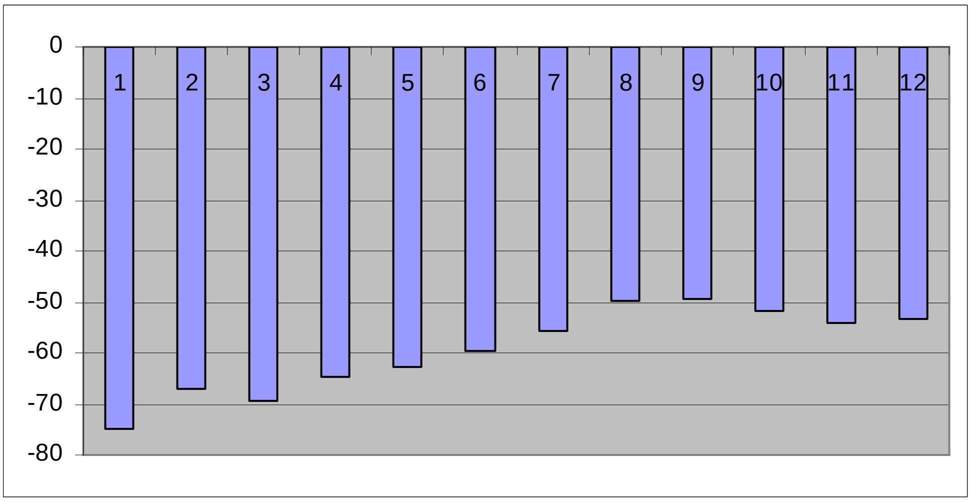


Figure 23 - Access point received signal strength variation across a corridor

Values higher than -10 dBm are only measured when the client is very close to an access point. The lowest value detected while performing these tests was close to -90 dBm. When an access point was not detected its RSSI value was assigned to -90 dBm, the lowest possible value to the location system.

$$R \in [0,-90]$$

While performing the NN training it was clear that this interval was not adequate. Due to the nature of the NN structure and hidden functions the input values must be normalized between lower intervals:

$$R \in [-1,1]$$

The values stored in dBm were normalized to the -1 and 1 interval:

$$R_{norm} = \frac{2(R - R_{\max})}{(R_{\max} - R_{\min})} - 1$$

Performing a standard average between readings of the same access point in the same point showed good results. However, there are more efficient methods to characterize the behavior of the RSSI at the same point. In order to minimize the effect of a high value of standard deviation of the samples taken from the calibration phase, a moving average was applied to the data set. The moving average has the capability of filtering out peak values that can influence the final average value. The moving average method did not improve the final error as demonstrated in the results section.

As mentioned before, more than one measurement is made at each point. Taking several measurements has the advantage of getting a clearer view of the signal behavior at that point. The number of measurements taken is also an important parameter for the location system both in calibration and location phase. In the calibration phase, since time is not a constraint, it is possible to acquire several readings to get a good training data for the NN. However in the location phase, time plays a crucial role. If the user requires a real-time location system, with minimum delay, the measurement number must be decreased at the cost of getting a lower precision. This way, the location system can be adapted to the user requirements. Higher precision involves a slower response from the location engine, since several measurements are being made. Despite this, typical location iteration with an average error of 2 meters takes between 1 and 2 seconds. The optimal interval between successive measurements was found to be between 300 and 400 ms. This value can also be changed, but due to the nature of the wireless hardware and software/driver implementation, lowering the sampling time has the effect of reporting the same RSSI. This way it is impossible to profile in an accurate manner the behavior of wireless radio signals. But since all the actors in the location system are connected to a network, there can be time constraints or other external sources of delays. Lowering the sampling time can mitigate these external factors, keeping the system speed at all times normal.

The NN number of outputs can be adjusted to the type of scenario. Most common applications only require a two dimensional output, consisting of a coordinate X and Y. This solution at most cases is sufficient and provides accurate results, however, a third coordinate can also be added to ensure a proper adaptation to multiple floor buildings.

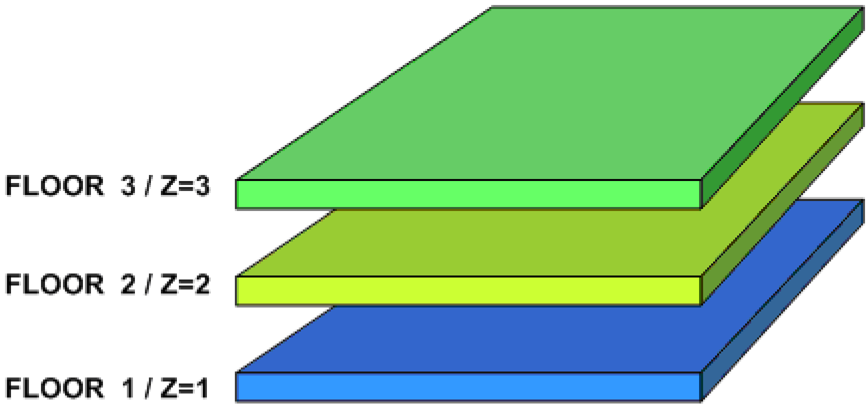


Figure 24 - Third dimension application scenario

The implication of adding a third dimension is an additional complexity to the NN structure. Since there are more points and also a third output, the number of hidden neurons must be raised in order that the NN can converge into a final solution.

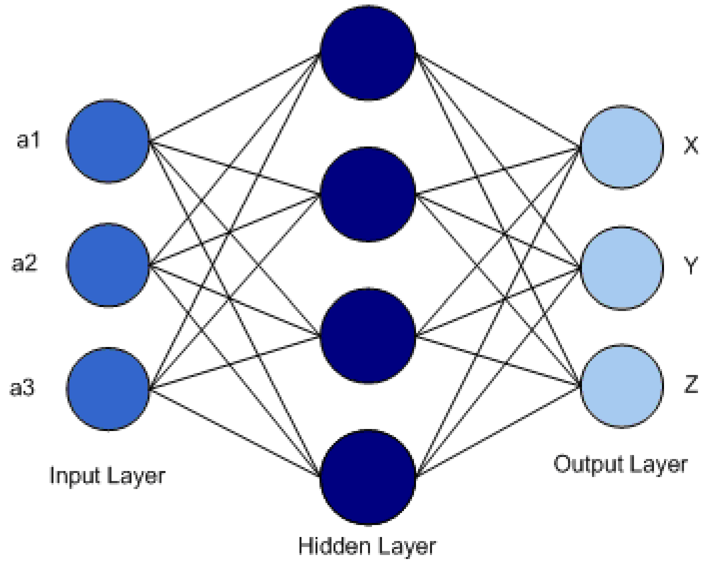


Figure 25 - Third output applied to neural network

The use of a third output makes the location system more capable of adapting to different scenarios, hence more scalable. Another approach was to divide the location area into different zones. These zones can be marked in different ways. One solution is to divide the location area with same size zones, typically squared or rectangular shaped.

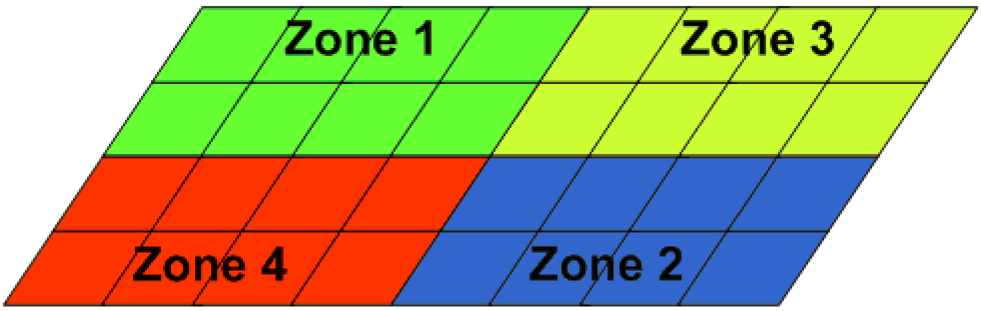


Figure 26 - Zone division implementation

This approach can use two or just one neural network. When using two neural networks, the first is used to calculate the zone where the client is located, or, in other words, which neural network should be used in the second iteration. The second neural network calculates the client position inside its defined zone. Considering the example of an area divided into four zones, the first neural network would output a number between one and four. Each zone has a unique trained NN, which will be run according to the output of the first NN. The neural network has no difficulty in finding the zone where the client is located. Still, there are some problems in the border areas. Another proposal for dividing the location area was to use the influence area of each access point. This is the area where a client associates or roams to an access point

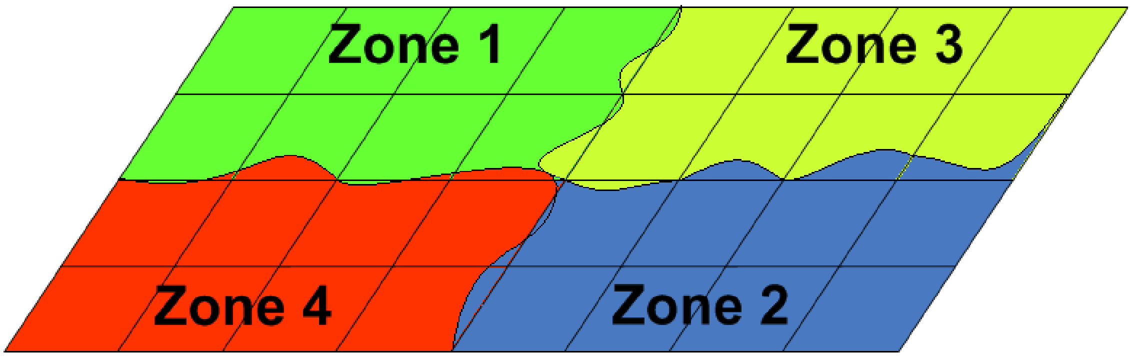


Figure 27 - Zone division based on access points area of influence

These solutions of dividing the area in zones presented good results, but similar to the ones provided by the less complex one-neural-network solution. It has the advantage of minimizing the peak error values in normal location iterations. However, sometimes some variation in the RSSI values can make the first NN to give an erroneous zone producing poor final output values. This is clearly a compromise between getting a stable system that is accurate most of the time and having high error values in abnormal operation mode. Using the simple method of one neural network, with no zones algorithm, has the advantage of making the location system more immune to abnormal RSSI values. This method, when combined with a good training and sufficient input data, can adapt to most of the situations.

Location system architecture

The location system architecture was developed keeping the initial requisites and goals in mind. The ideal solution was to have a location system capable of tracking clients seamlessly. The system core is constituted by two main parts: the server, which processes requests and provides the final location information and the client software, which collects RSSI values and it is configured to communicate easily with the server.

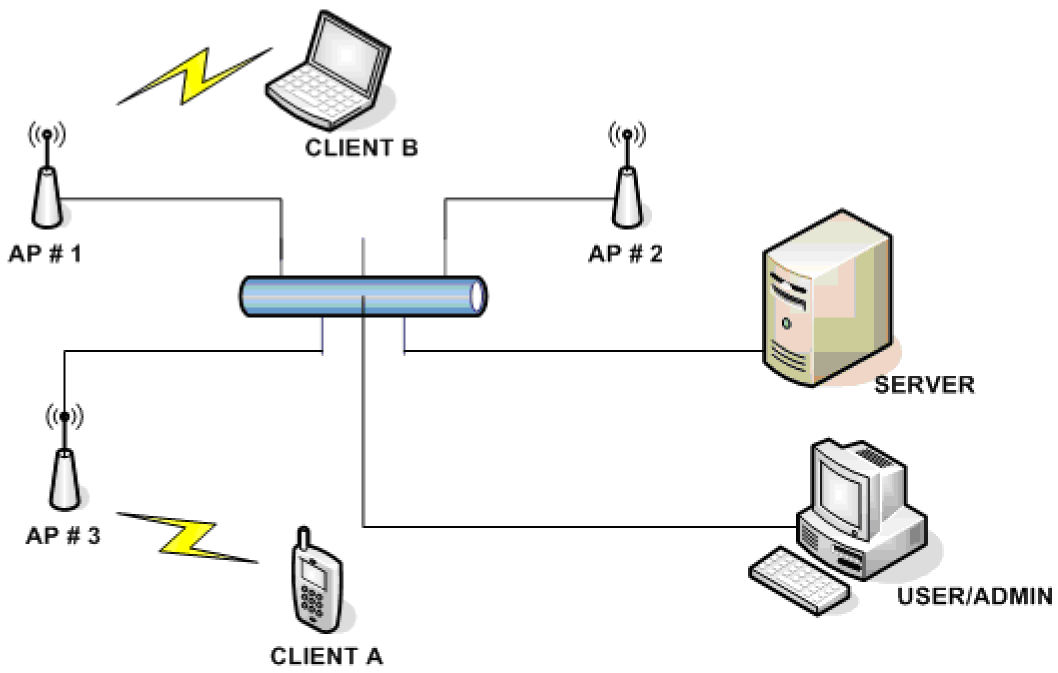


Figure 28 - Location system architecture

The server has two types of services installed. The first service is responsible for collecting RSSI values from all the existing wireless clients. It also has the responsibility of calculating the client position using the NN algorithms. This service interacts with a second one, a web interface, which was developed to allow an easy access to the location system. With this web interface a user can ask the server to locate a specific wireless client and display that information on a map or a floor plan.

The location process is quite simple but all the actors play a crucial role in all its steps. First, a location request must be sent to the server using the web interface. This request has all the information that the server needs:

- Client's IP Address
- Sampling period (default value: 300 ms)
- Number of samples per location (default value: 5)

The sampling period is the time interval between consecutive measurements made by the wireless client. The number of samples per location is the number of measurements made by the client before calculating its position. A high value on this setting improves the location system accuracy but also introduces a high time delay.

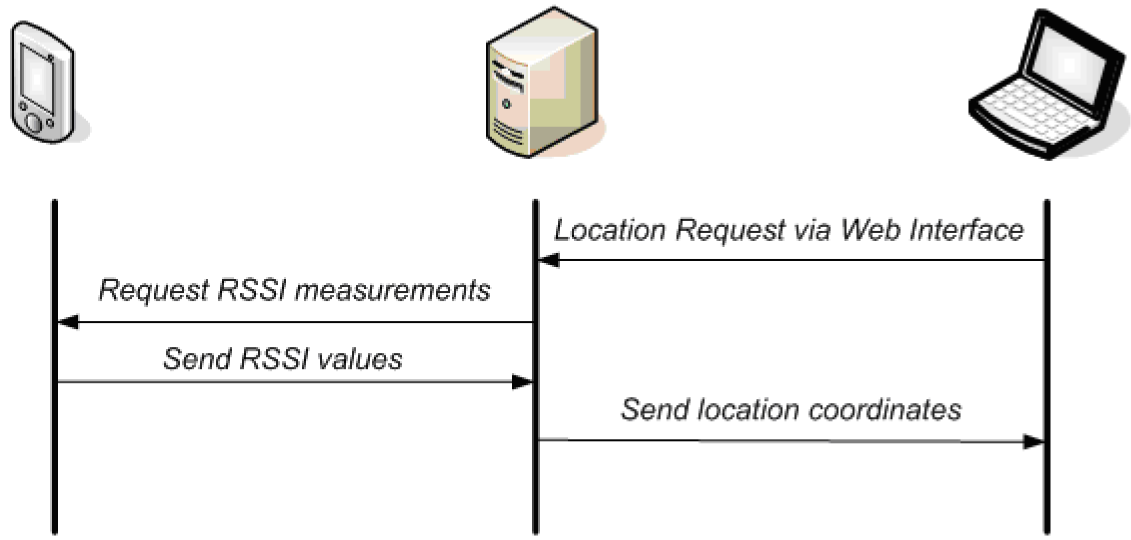


Figure 29 - Location process iterations

After receiving the location request, the server asks for RSSI measurements from the wireless client. The existing client software processes the request and retrieves RSSI values from the wireless card and sends them back to the server. It is important to note that the client software reports RSSI values from all the access points detected. The software is constructed to minimize the client's awareness of the location system. It has minimal memory occupancy and processor usage. The bandwidth used is also kept at low values. The message sent from the server, asking the wireless client to perform signal strength measurements, is repeated until the number of samples chosen in the web interface is reached.

After collecting the measurements sent by the client the server filters out unused access points by making a comparison to the ones used during the calibration phase and calculates the wireless client position. This information is sent back to the users that request the location displayed on its screen.

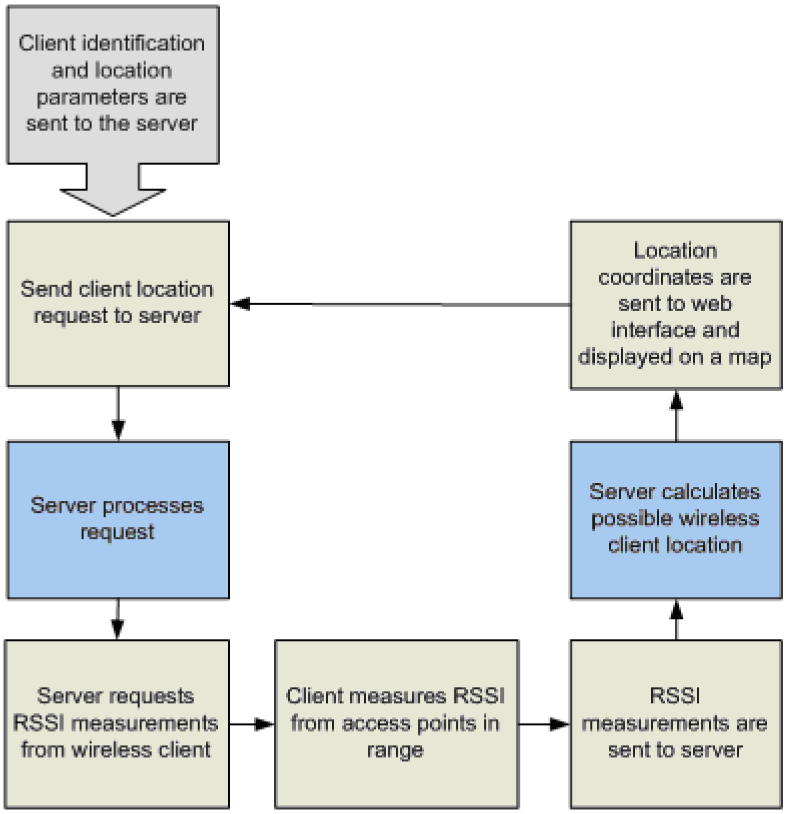


Figure 30 - Location system algorithm

All these steps can take last from one to two seconds. This time, however, depends on both the sampling time chosen and the quality of the connection link between the three system parts (user, server and wireless client).

The sampling time used (default value is 300 ms) and the number of samples per location iteration can be changed and can be adapted to the user's accuracy demands as mentioned before.

Software solution

The first client software version was developed in Linux. It allowed sufficient flexibility for a project of this type and provided a clear insight of the problems that exist when the objective is to acquire RSSI values from wireless cards. Despite the growing number of special developed Linux wireless drivers there is not a standard way of calculating the reported RSSI values from the wireless card. To make things worst, many drivers report erroneous values and the speed of scanning all the wireless channels change from card to card. Linux is also not a main choice for operative system for most people, and the location system should be scalable and adapted to most situations. To solve all these problems, a Windows software solution was developed. This software relies on NDIS [27], a special intermediate protocol between the application level and network hardware.

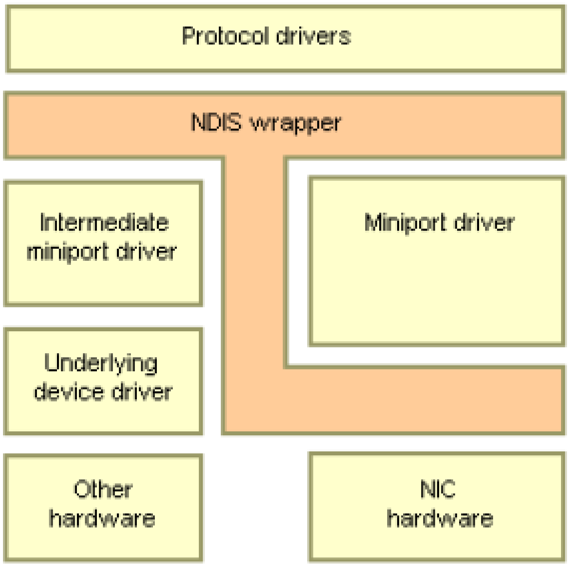


Figure 31 - NDIS architecture

NDIS provides a standard that the client software can rely, since Microsoft has special rules that wireless card drivers must follow. This way the client software can retrieve RSSI values from NDIS 802.11 data fields. However, Windows does not keep real time RSSI values and it only updates them at large time intervals (typically 30 seconds). Clearly, there was the need of a solution capable of performing RSSI values updates at each request. Fortunately, besides reporting values, NDIS also allows hardware iteration indirectly so, it was possible to request the wireless card to perform scans on demand, therefore reporting real time RSSI values [28].

The web interface developed allows a user to visualize the client's location in a floor plan or map. This image can be customized to the location scenario. There were two versions of this web interface. The main difference between each one is the technology used.

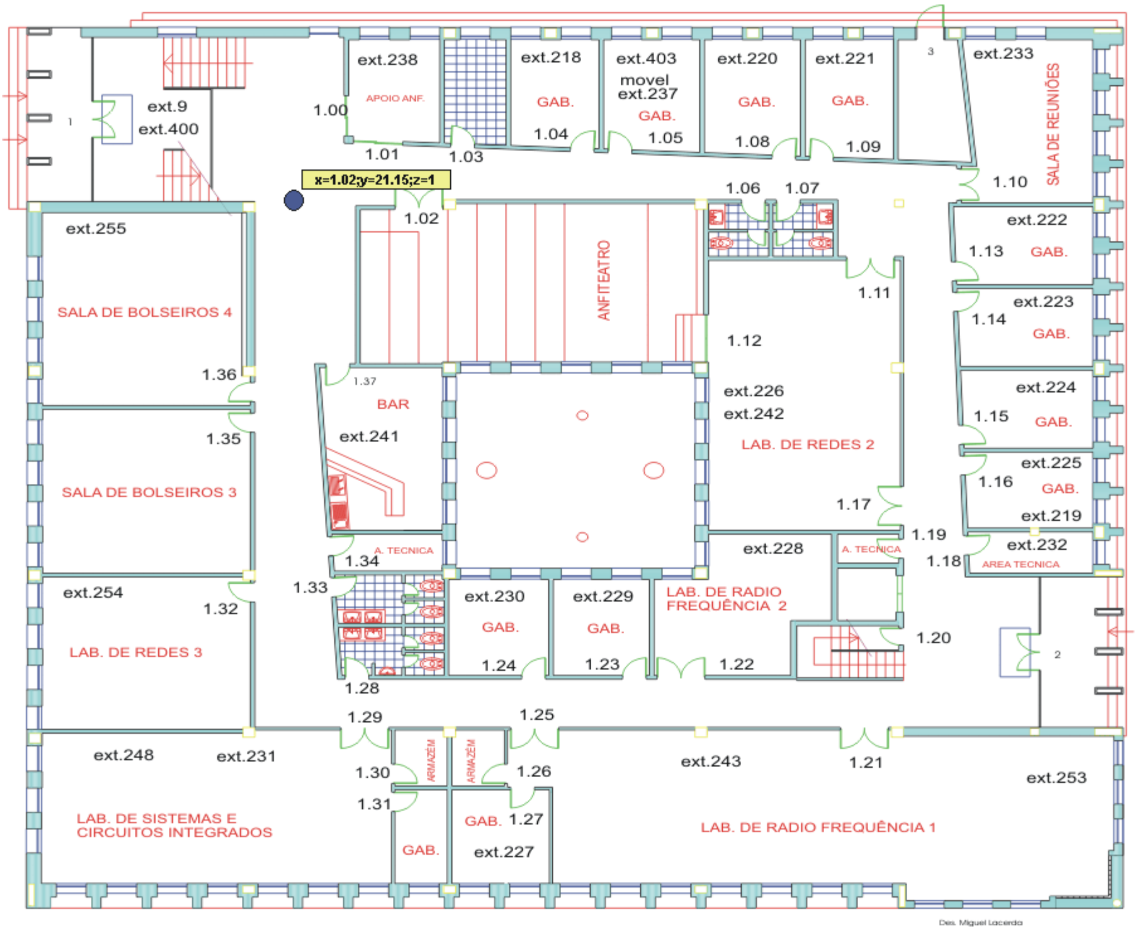


Figure 32 - Client side program: AJAX implementation

The first version, and most reliable, is a Java applet [29] that communicates directly to the server and keeps location information up to date at real time. The second version is an AJAX [30] implementation of the Java applet. It has the advantage of not requiring the user to install Java on its machine web browser.

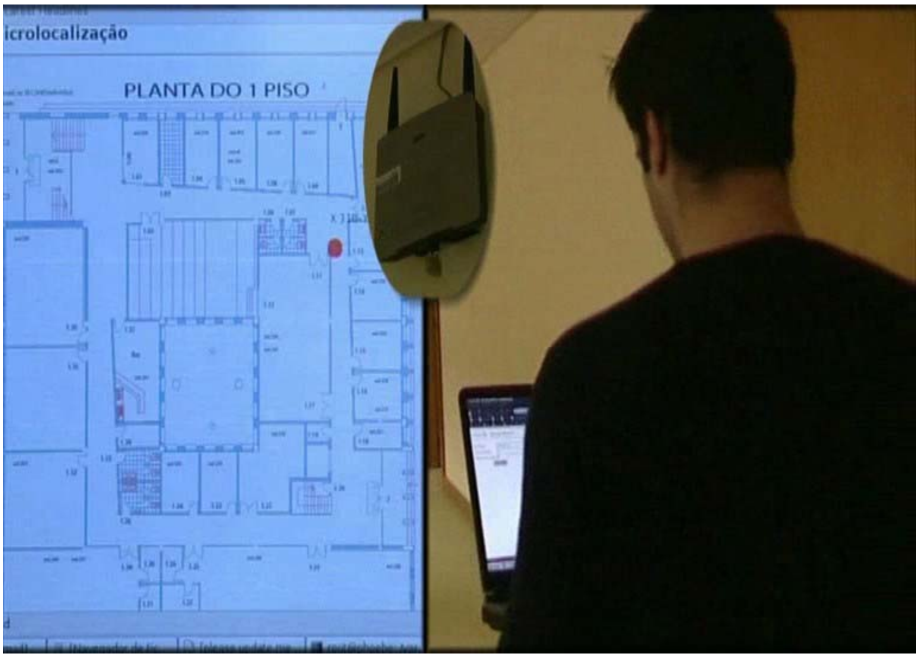


Figure 33 - Location system in action (Java implementation)

Both of the versions display a real time location map with a clear indication of the user position.

8 Results

RSSI change over time

Since the location system performance depends on the RSSI values from the various access points, it is important to study the behavior of these signals over time. The standard deviation from the several RSSI values at the same location can not take large values, since it will degrade the location system performance. Normally, the transmitted power in access points in a typical WLAN is constant. The results presented here were collected across a corridor in the *Instituto de Telecomunicações* main building.

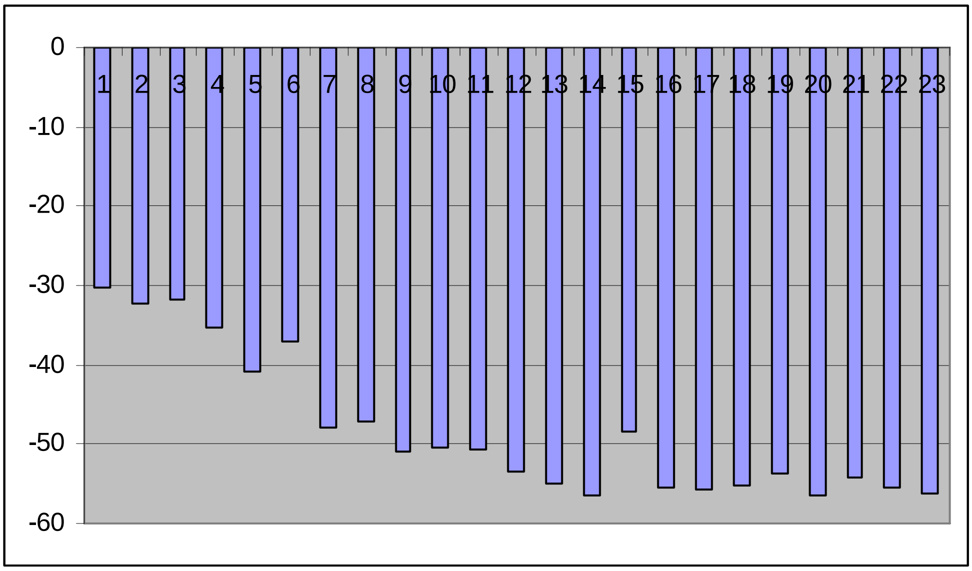


Figure 34 - Average RSSI values measured across IT's South corridor - October 2005

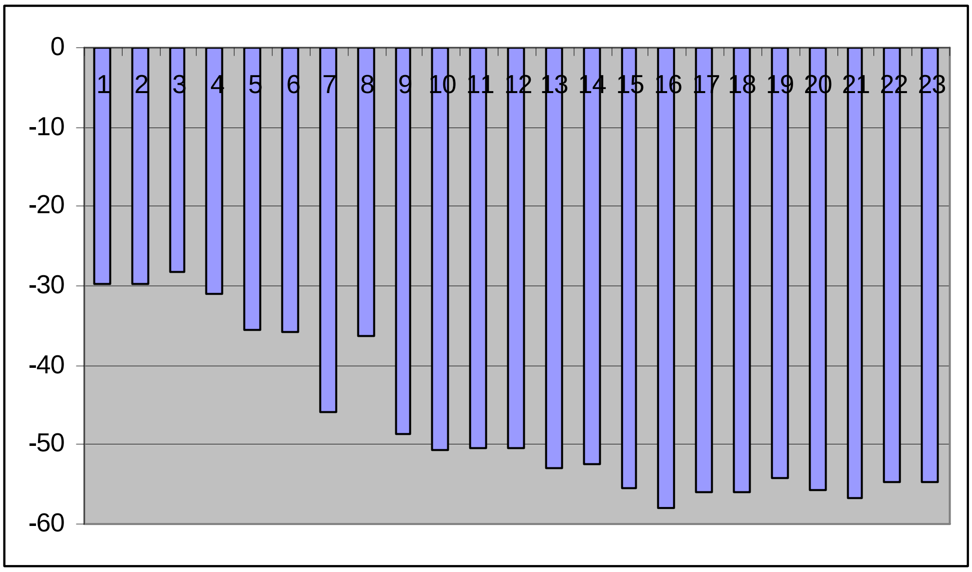


Figure 35 - Average RSSI values measured across IT's South corridor - November 2005

These results show that, despite the changes in the RSSI values measured, the standard variation does not take large values. In order to get a wider view of a typical location system implementation scenario, several tests were made in a shopping centre, where the radio propagation environment is typically harsh since there is a constant movement of people. This test bench had four day duration and consisted of placing a WiFi device collecting 20 measurements at each time in intervals of five minutes. The access points used on this test were the ones that existed on the shopping centre that are from several ISPs. Since the shopping centre entrances are equipped with counting sensors, the RSSI results were crossed with the number of people that were inside the building. This way it might be possible to conclude something about the effect of the constant movement of people and its effect on the radio propagation.

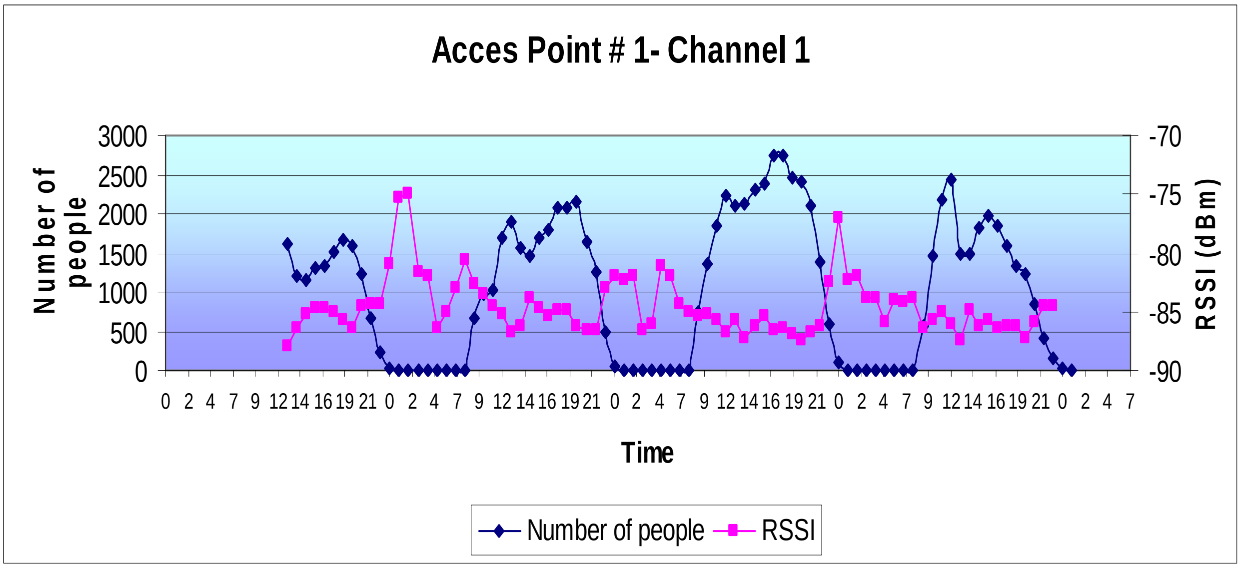


Figure 36 - Shopping center scenario - RSSI variation over time and visitor number influence – channel 1

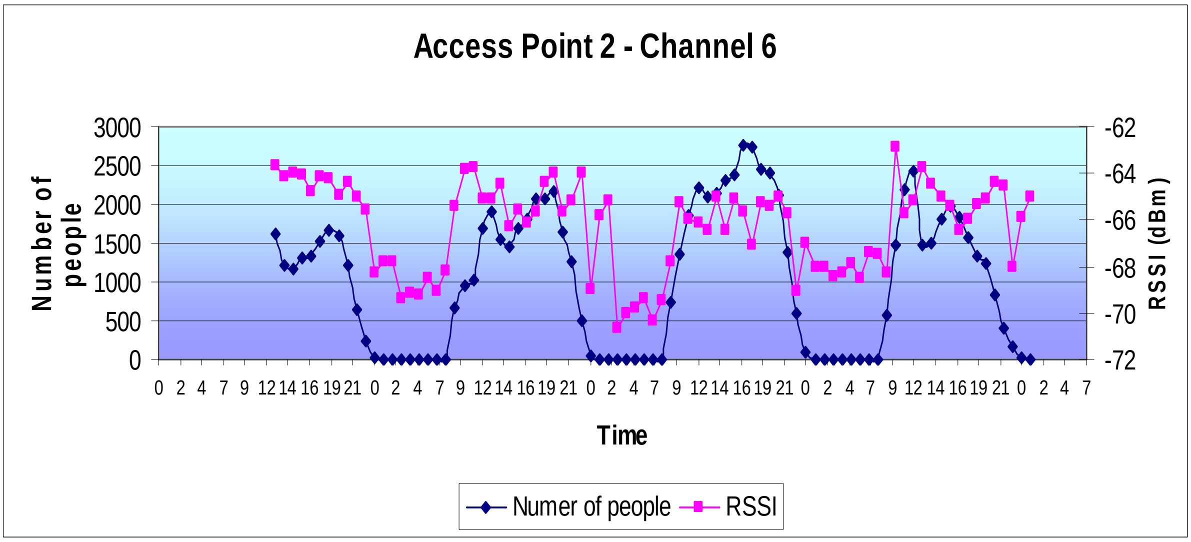


Figure 37 - Shopping center scenario - RSSI variation over time and visitor number influence – channel 6

Analyzing the results, it is clear that the number of people roaming in the shopping centre has an effect of the RSSI value measured. But, most of the time, the RSSI variation is not large enough to have a profound impact on the location system. A signal variation between 0 and 4 dB is a typical small scale fading value, which is observed on the measurements taken at the same point with 300 milliseconds intervals.

ANN parameters values effect

The behavior of the artificial neural network, as mentioned before, has a dependency on the parameters values. The mean square error (MSE) value provides an understanding of the number of neurons that should be used.

Crossing the MSE values with the average error gives the optimal number of neurons that, in this case, is near twelve.

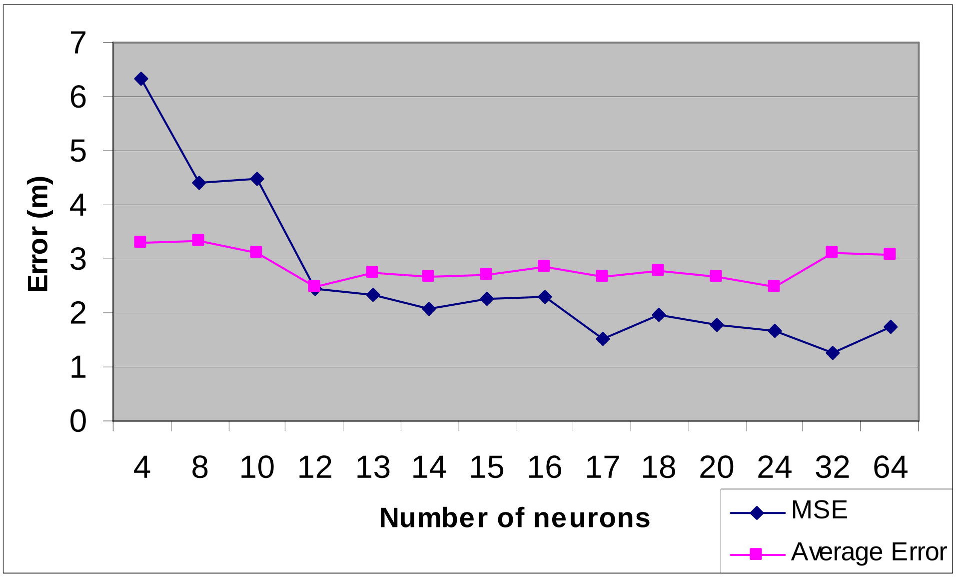


Figure 38 - Average location and training error versus number of neurons

In this situation the ideal number of neurons of the hidden layer was found to be twelve. This number makes a good commitment between the training MSE and the average error. Nevertheless, the number of neurons used in the hidden layer must be adapted to each situation and to the size of the training data set.

During the location phase the wireless client is asked to perform RSSI measurements. The number of measurements has influence in the accuracy and duration of the location process. Increasing the number of measurements improves accuracy but increases the location duration. A good compromise between location response time and accuracy is a number that goes from five and eight readings.

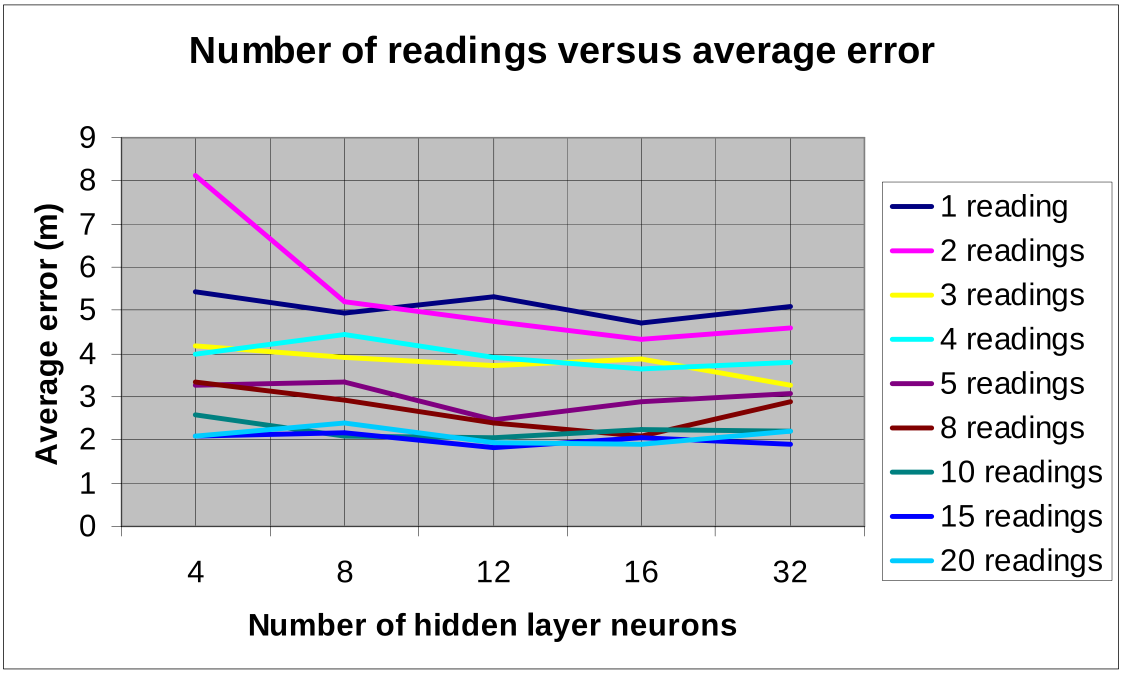


Figure 39 - Number of readings influence on the average location error

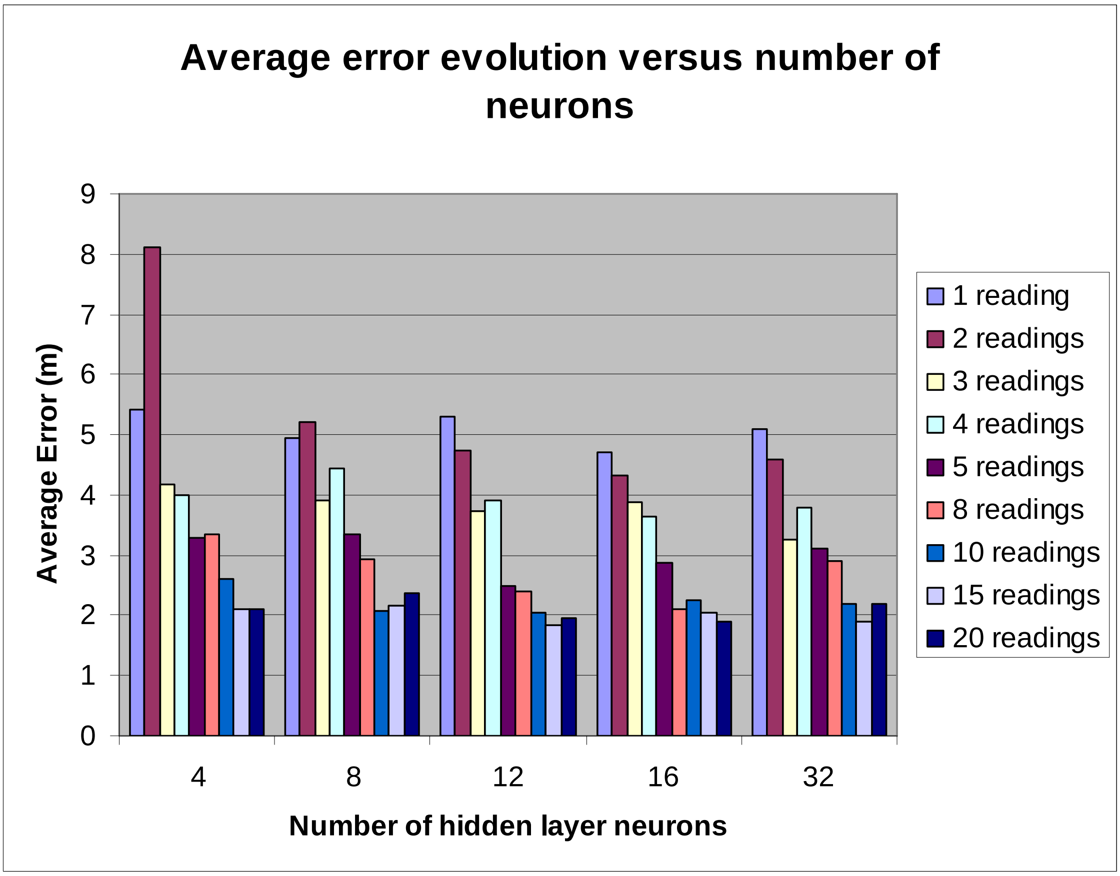


Figure 40 - Number of neurons and readings influence on the average location error

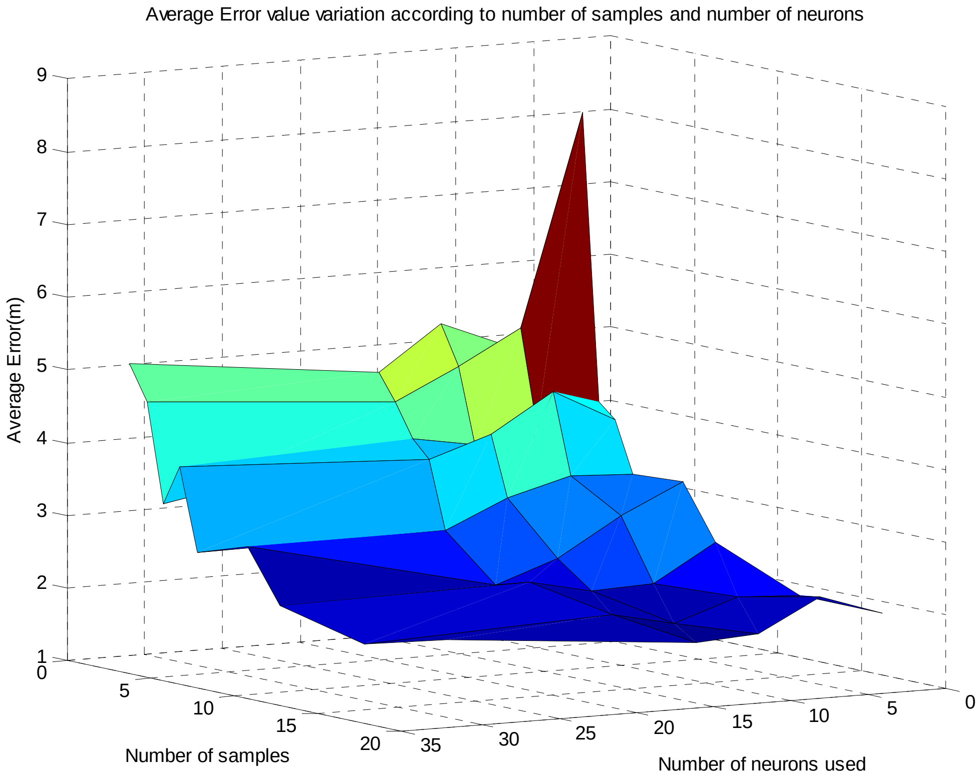


Figure 41- Average Error variation versus number of neurons and sampling size

Analyzing the behavior of the average error when there is a variation in the number of neurons in the ANN hidden layer and the number of samples (readings) used, it is possible o understand that there is an optimal neural network size. According to figure 42, the average error is minimized when the number of neurons of the neural network hidden layer is between 10 and 16. The sampling size also matters, as it is shown on the graphic, a low number of samples, has a negative impact on the location system performance. A low number of samples degrades the final average error values, compromising the system accuracy. This phenomena is explained when one thinks of radio propagation characteristics. Indoor propagation is always affected by effects like small scale fading, multipath, scattering and diffraction. Retrieving just one sample is just too low, since the location algorithm is based on the average RSSI values.

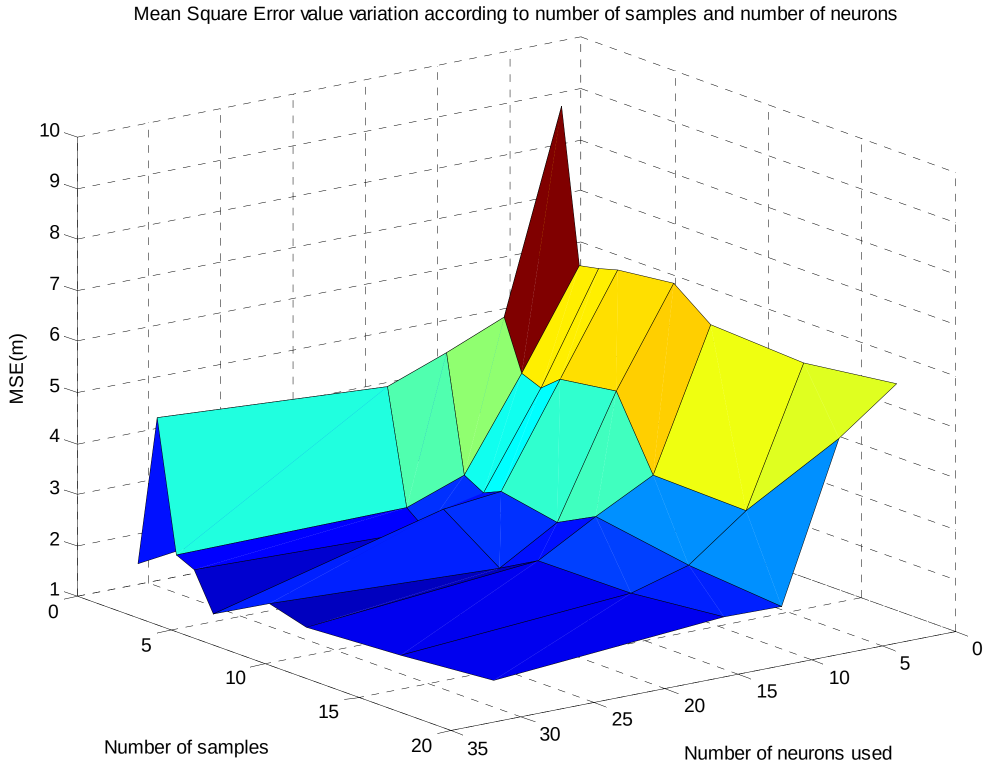


Figure 42 - MSE variation according to number of neurons and samples

The mean square error is the output value of the neural network training. It is a figure of merit of the NN training and according to its value it is possible to conclude how well the NN has adapted to the input values. As figure 40 shows, MSE has large values when a low number of neurons is used. This is a natural behavior of the neural network, since there is a large set of data, and it is almost impossible to converge to a good final solution with a low number of neurons no matter the number of readings. A low number of readings also degrades the neural network capability of converging. The adequate number of readings should be higher than five. A higher number of readings does not significantly improve the neural network learning performance. Clearly, the number of neurons used has a high impact on the mean square error of the learning phase.

Average Error

It is important to study the behavior of the error in the location system. A low average error is not always a sign of optimal performance. The error histogram provides a clear insight of the location system accuracy.

Using different settings, like the number of hidden neurons, access points and readings, it is possible to acquire the optimal settings for different scenarios and user requirements.

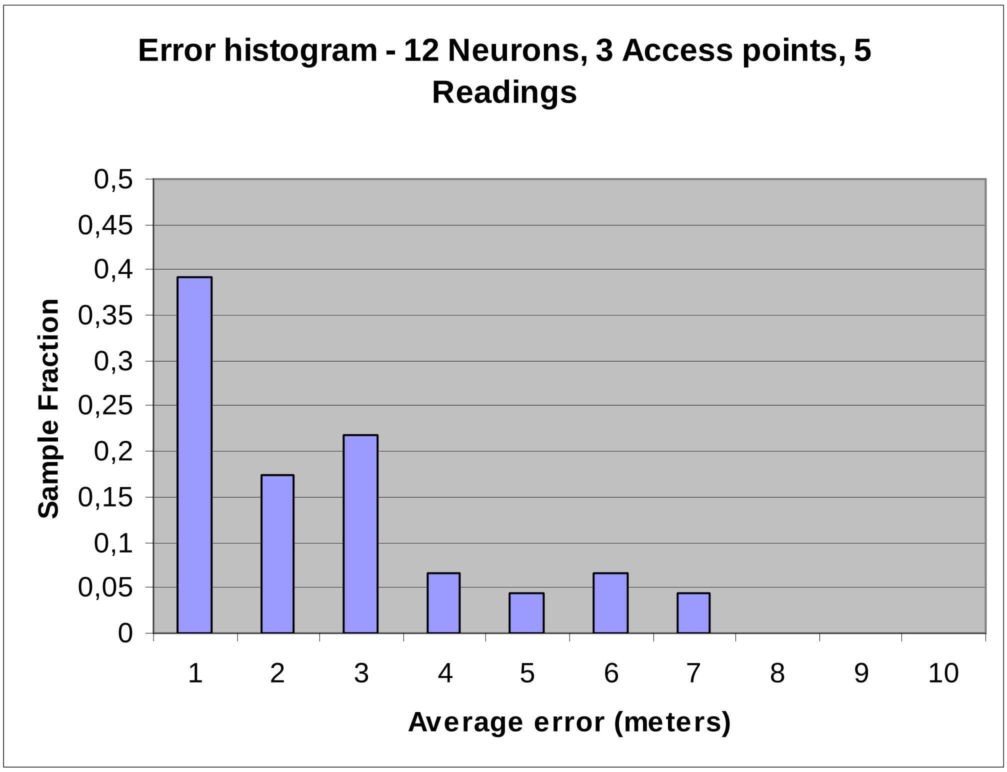


Figure 43 - Error histogram using 12 neurons, 3 access points and 5 readings

Using a sufficient number of readings for an adequate accuracy (5 readings), the average error value is 2.4 meters. The maximum error value is 7 meters, and for about 80% of the samples, the error is below 3 meters. This result is improved increasing the number of readings to 20, where the average error drops down to 1.9 meters.

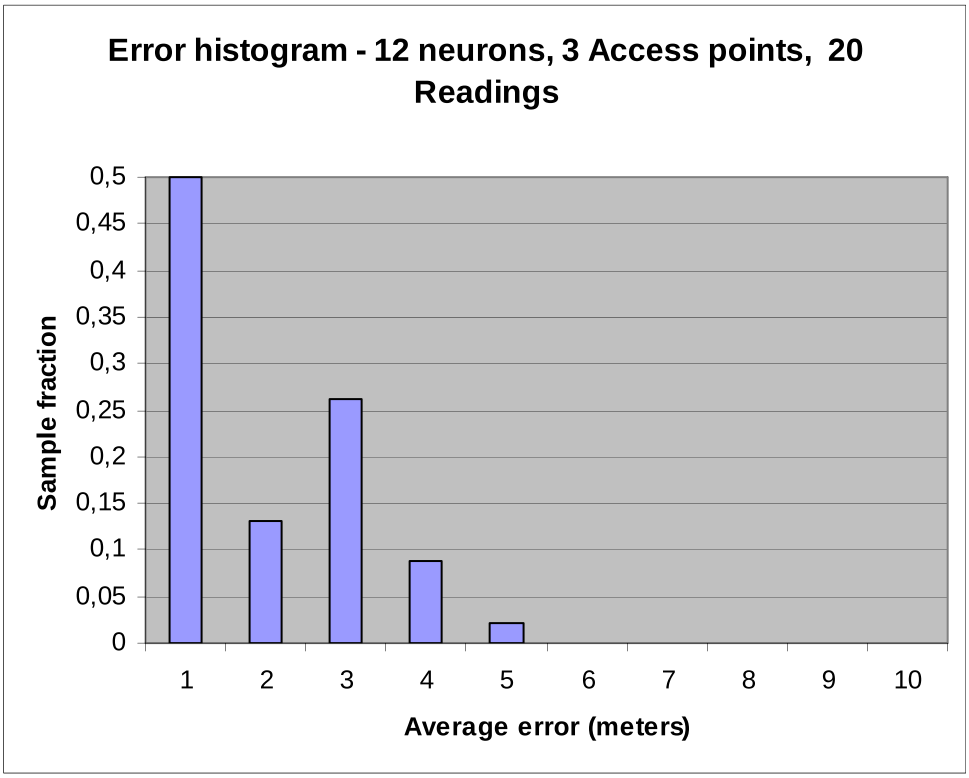


Figure 44 - Error histogram using 12 neurons, 3 access points and 20 readings

Using additional measurements decreases the maximum error to five meters. According to figure 43 the location system offers an error less than one meter for 50% of the samples. Also for 90% of the samples the error is below three meters.

9 Conclusion and future work

This work provides a clear description of a wireless location system. Using a simple software solution, with no additional hardware, the location system has a good accuracy that can be improved by increasing the number of readings at the cost of lowering the location response time.

The use of RF fingerprinting is a valid solution to provide accurate positioning techniques. Nevertheless, it can be time consuming gathering measurements and profiling a large indoor scenario. This time was minimized using a simple and fast RSSI measurement and calibration tool, which can be used in an easy way. The usage of radio frequency fingerprinting requires employing mathematical approaches to solve the location problem. Methods like using propagation models and nearest neighbor were evaluated and it was concluded that their performance was inadequate. Artificial neural networks applied to the location paradigm offer sufficient adaptability between different scenarios, contrary to other algorithms used in other location applications. ANN based location algorithm provides enough flexibility and its final accuracy competes directly with the best known location applications.

The idea behind an indoor location system is its capability of providing accuracy while being simple to use. The location system developed can be easily deployed in existing WLAN with minimal cost and difficulty. Additionally it does not require any changes in the existing WLAN infrastructure, only a small software program must be installed in the client's device. This software does not interfere with the client's normal usage of its device. The system robustness is not compromised by minimal environmental changes in the indoor scenario, and it is immune to the multi-path and small scale fading effects, typical encountered on indoor radio propagation.

This location system is divided into two parts, the macro and the micro-location. Although some testing was made to ensure the interaction between those two systems, a final and efficient solution was not achieved. The main goal is to choose the adequate neural network according to the output of the macro-location system. This approach is similar to the one used in the zones algorithm. Possible future work includes a seamless integration between the two systems.

The hidden function of an artificial neural network is one of its core elements. The development of hidden function dedicated to the behavior of wireless signals propagating in an indoor scenario, can be of the future goals of this work.

Presently, the local system has been deployed in two different scenarios with immediate success. In the future, one can expect to improve its accuracy and lower the time required to perform the initial calibration.

10 References

1 J. J. Caffery, Wireless Location in CDMA Cellular Radio Systems. Norwell, EUA: Kluwer, 1999.

2 A. Ward, A. Jones, and A. Hopper. A new location technique for the active office. IEEE Personal Communications,4(5):42–47, Oct. 1997.

3 N. B. Priyantha, et al, The Cricket Location-Support System, Proc. 6th Ann. Int'l Conf. Mobile Computing and Networking (MOBICOM 2000), ACM Press, New York, 2000, pp. 32-43

4 R. Want et al., The Active Badge Location System, ACM Trans. Information Systems, Jan 1992, pp 91-102

5 Jay Werb and Colin Lanzl. Designing a positioning system for finding things and people indoors. IEEE Spectrum, September 1998 pp 71–78

6 J. Hightower, et al. SpotON: An indoor 3D location sensing technology based on RF signal strength.Technical Report UW CSE 00-02-02, Department of Computer Science and Engineering, University of Washington, Seattle, WA, Feb. 2000.

7 Orr, Robert, et al.. The Smart Floor: a Mechanism for Natural User Identification and Tracking. Human Factors in Computing Systems, 2000, Georgia Institute of Technology.

8 Location Technologies for GSM, GPRS and UMTS Networks. SnapTrack. 2003.

9 Bahl, P. and Padmanabhan, V.N.. (2000). RADAR: An in-building RF-based user location and tracking system. In: EEE Infocom 2000, Tel Aviv, Israel, March 2000.

10 Roberto Battiti, et al. Location-aware computing: a neural network model for determining location in wireless lans. Technical Report DIT-5, Universit`a di Trento, Dipartimento di Informatica e Telecomunicazioni, 2002.

11 Kaemarungsi, Kamol, et al. Modeling of Indoor Positioning Systems Based on Location Fingerprinting. IEEE, 2004, School of Information Science. University of Pittsburgh.

12 Rodriguez, Marcela D., et al. Location-Aware Access to Hospital. IEEE Transaction on Information Technology in Biomedicine, Dec. 2004.

13 Madigan, David, et al. Location Estimation in Wireless Networks. A. Rutgers University. 2005.

14 Haeberlen, Andreas, et al.. Practical Robust Localization Over Large-Scale 802.11. MobiCom'04, 2004, Rice University. Philadelphia: ACM, 2004.

15 P. Krishnan, et al, "A System for LEASE: Location Estimation Assisted by Stationary Emitters for Indoor RF Wireless Networks," IEEE Infocom, Hong Kong, March 2004.

16 "Ekahau Positioning Engine." Ekahau. <http://www.ekahau.com>.

17 Kontkanen, Petri, et al. Probabilistic Methods for Location Estimation in Wireless Networks. Helsinki Institute for Information Technology HIIT. Helsinki: HIIT Technical Reports, 2004.

18 Vossiek, Martin, et al. Wireless Location - Concepts, Solutions and Applications. Siemens Corporate Technology,. Munich.

19 "Aeroscout - Enterprise Visibility Solutions." Aeroscout. <http://www.aeroscout.com/>.

20 Wi-Fi Location-Based Services. Cisco. San Jose: Cisco, 2006.

21 Rappaport, Theodore. Wireless Communications: Principles and Practice. Vol.2. Prentice Hall PTR, 2001.

22 “A Simple Network Management Protocol (SNMP) RFC 1157." IETF. <http://tools.ietf.org/html/rfc1157>.

23 "BSD Syslog Protocol - RFC 3164." IETF. <http://tools.ietf.org/html/rfc3164>.

24 M.T. Hagan et al., Neural Network Design. Boston, EUA: PWS, 1996

25 "Fast Artificial Neural Network Library." FANN. <http://leenissen.dk/fann/>

26 Recurrent Neural Networks Design and Applications. London: CRC Press, 2001.

27 "NDIS - Network Driver Interface Specification." Windows Hardware Developer Central. <http://www.microsoft.com/whdc/device/network/ndis>.

28 Divine, Thomas F. "PCAUSA - Rawether for Windows." PCAUSA. <http://www.rawether.net>.

29 "Java Technology." Sun Developer Network. <http://java.sun.com>.

30 "W3C Technical Reports and Publications." W3C. <http://www.w3.org>.