ENVIRONMENT AWARE LOCATION ESTIMATION IN CELLULAR NETWORKS

by

Onur Türkyılmaz

B.S., Computer Engineering, Yeditepe University, 2005

Submitted to the Institute of Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2007

ENVIRONMENT AWARE LOCATION ESTIMATION
IN CELLULAR NETWORKS

APPROVED BY:

Assoc. Prof. Fatih Alagöz (Supervisor)	.....
Prof. Emin Anarım	.....
Assist. Prof. Tuna Tuğcu	.....

DATE OF APPROVAL: 13.09.2007

ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Prof. Fatih Alagöz for his invaluable guidance and support not only throughout this M.S. study but also during my entire graduate education. I also thank Assoc. Prof. Tuna Tuğcu for his precious observations and advices. I am also grateful to Prof. Emin Anarım for his encouragement and support in my graduate education. I would also like to thank all of my colleagues and friends in SATLAB for creating such a cheerful environment to work in. Furthermore, I appreciate TÜBİTAK for their financial support during my graduate education.

I would like to thank my friend Gürkan Gür for his great help and friendship whenever I needed throughout this study.

Finally I thank my family for being the greatest support during my whole educational life.

This work was supported by the State Planning Organization of Turkey under the Next Generation Satellite Networks Project, DPT 03K 120250.

ABSTRACT

ENVIRONMENT AWARE LOCATION ESTIMATION IN CELLULAR NETWORKS

Location Based Services (LBS) enable personalized services to the mobile subscribers based on their current position and consequently it has received significant attention in both research and industry over the past few years. Mobile positioning plays a key role in providing LBS such as wireless emergency services, location tracking services and location-aware information and advertisement services.

Using received signal strength (RSS) measurements from the control channels of several base stations, the location of a mobile unit can be estimated. Although the RSS method is not as precise as other localization methods in literature such as angle of arrival, time of arrival, and assisted global positioning system, it is easy to implement on any cellular network as it does not require any changes to existing phones and network structure. Since radio propagation characteristics vary in different environments, knowing the environment of the mobile user is essential for accurate RSS based location estimation. In this study, a novel mobile positioning algorithm for cellular networks based on the estimation of the radio propagation environment is presented. The key feature of the proposed method is its capability to estimate the environment of the mobile user as urban, suburban or rural using pattern recognition and to utilize this information for enhancing RSS based distance calculations. The proposed algorithm has been evaluated using field measurements collected from a GSM network in diverse geographic locations. Our approach turns out to be significantly beneficial, enhancing estimation accuracy, and thereby enabling high-performance mobile positioning in a practical and cost effective manner. Additionally, it is computationally light-weight and can be integrated onto any received signal strength based algorithm as an enhancement add-on.

ÖZET

HÜCRESEL AĞLARDA ORTAM-BİLİNÇLİ KONUM BELİRLEME

Konuma dayalı servisler, mobil kullanıcılara bulundukları konuma bağlı olarak kişisel servisler sunmayı mümkün kılmaktadırlar ve dolayısıyla bu durum özellikle son yıllarda telsiz ağlarda konum belirleme üzerine çalışmaların artmasına neden olmuştur. Konum belirleme sayesinde acil durum servisleri, iz sürme, konuma dayalı bilgi ve reklam servisleri, konuma dayalı ücretlendirme gibi birçok servis mümkün hale gelmektedir.

Bir mobil kullanıcının konumu birkaç baz istasyonunun kontrol kanallarından gelen sinyal gücü ölçümlerini kullanarak tahmin edilebilir. Sinyal gücü ölçümlerine dayalı bu yöntem literatürdeki sinyal geliş açısı yöntemi, sinyal alım zamanı yöntemi ve yardımlı küresel konumlama sistemi gibi yöntemler kadar hassas olmasa da, ağ yapısında ve mobil cihazlarda herhangi bir değişiklik ve buna bağlı olarak ek maliyet gerektirmemesinden dolayı yaygın olarak kullanılmaya devam edilmektedir. Radyo sinyallerinin yayılma karakteristiği ortamdan ortama farklılıklar gösterdiği için, konumu belirlenecek olan mobil cihazın içinde bulunduğu ortam koşulunun tahmini, sinyal gücüne dayalı konum belirleme yöntemlerine büyük yarar sağlayarak hassasiyeti arttıracaktır. Bu çalışmada, hücresel ağlar için ortam-bilinçli, sinyal gücü ölçümlerine dayalı yeni bir konum belirleme yöntemi sunulmaktadır. Önerilen metodun ana özelliği, örüntü tanıma vasıtasıyla kullanıcının içinde bulunduğu ortamı kırsal kesim, banliyö veya şehir içi olarak belirleyip bu bilgiyi kullanarak sinyal gücüne dayalı konum belirleme metodunun hassasiyetini arttırmaktır. Yöntemin başarımı farklı coğrafi koşullarda bir GSM ağından alınan gerçek ölçümler kullanılarak test edilmiş ve konumlama başarımının ortalama hata ve standard sapma gibi birçok yönden iyileştiği gözlemlenmiştir. Bunlara ek olarak, sunulan metod herhangi bir sinyal gücü tabanlı konum belirleme algoritmasına bir iyileştirme eklentisi olarak kolayca eklenebilir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZET v
LIST OF SYMBOLS/ABBREVIATIONS viii
LIST OF FIGURES xi
1. INTRODUCTION 1
2. RADIO WAVE PROPAGATION 4
2.1. Basic Principles 4
2.1.1. Free-space Attenuation 6
2.1.2. Absorption 7
2.1.3. Reflection 7
2.1.4. Diffraction 9
2.2. Propagation Models 10
2.2.1. Hata Model 11
2.2.2. Walfisch-Bertoni Model 12
3. MOBILE POSITIONING TECHNIQUES 15
3.1. Cell Global Identification and Timing Advance 16
3.2. Signal Strength 17
3.3. Time of Arrival 21
3.4. Time Difference of Arrival 24
3.5. Angle of Arrival 26
3.6. Assisted GPS 28
3.7. Database Correlation 29
4. MACHINE LEARNING METHODS USED IN EARBALE 31
4.1. The Artificial Neural Network 31
4.1.1. Multilayer Perceptron 33
4.1.2. Back-propagation Algorithm 34
4.2. The Decision Tree 38
4.2.1. Decision Tree Building 39

4.2.1.1. Splitting Node 40
4.2.1.2. Stopping Criterion 40
4.2.1.3. Tree Prunning 41
4.2.1.4. Cross Validation 41
4.2.1.5. Finding the Optimal Tree Size 42
5. ENVIRONMENT AWARE LOCATION ESTIMATION (EARBALE) 43
5.1. Data Collection and Preparation Phase 44
5.2. ANN-based Classification Phase 46
5.2.1. Dimensionality Reduction Subphase 47
5.2.2. Training Subphase 48
5.3. Estimation (Triangulation) Phase 50
5.4. Evaluation Phase 52
6. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION 54
7. CONCLUSIONS 56
REFERENCES 58

LIST OF SYMBOLS/ABBREVIATIONS

ΔV_{ij}	Weight correction term
a	ANN output vector
$a(h_m)$	Correction factor for mobile unit antenna height
b	Street width
b	ANN bias vector
c	Speed of light
d	Distance to the antenna
E	Electric field
F	Power density flow
f	Carrier frequency
G_r	Gain of the receiving antenna
G_T	Gain of the transmitting antenna
h_b	Building height
h_m	MS antenna height
h_t	Transmitting station antenna height
I	Current
I_0	Current amplitude
k	Phase constant
L	MS antenna height
lat_e	Latitude of the estimated point
lat_{GPS}	Real latitude of the MS
lat_i	Latitude of the i^{th} of n intersection points
L_f	Free-space loss
L_{ms}	Multiscreen loss
lon_e	Longitude of the estimated point
lon_{GPS}	Real longitude of MS
lon_i	Longitude of the i^{th} of n intersection points
L_{rts}	Roof-to-street diffraction and scatter loss
n	Channel noise
o	ANN output
PL	Path loss
P_R	The received power
P_T	Transmitted power

Q	Minimum split criterion value
r	Distance between MS and BTS
r_{ij}	Difference in time delays between the MS and BTSs i and j
Rs	Radiation resistance of MS antenna
t^0	Time instant that the burst signal is transmitted
t_i	Time of arrival of the received signal
u	Coordinate unit vector
w	Frequency
W	ANN weight matrix
α	Signal fading
γ	Absorbtion constant
δ_j	Error information term
ε	Error
θ	Phase angle
λ	Wavelength

A-GPS	Assisted GPS
ANN	Artificial Neural Network
AOA	Angle of Arrival
BSC	Base Station Controller
BTS	Base Transceiver Station
CGI	Cell Global Identification
DCM	Database Correlation Method
DT	Decision Tree
EARBALE	Environment-Aware RSS-Based Location Estimation
GMLC	Gateway Mobile Location Center
GPS	Global Positioning System
GSM	Global System for Mobile Communications
LBS	Location-Based Services
LMU	Location Measurement Unit
LOS	Line of Sight
MLP	Multi-Layer Perceptron
ML	Maximum Likelihood
MS	Mobile Station

MSC	Mobile Switching Center
MSE	Mean Square Error
NLOS	Non-Line-of-Sight
RSS	Received Signal Strength
SMLC	Serving Mobile Location Center
TA	Timing Advance
TDOA	Time Difference of Arrival
TOA	Time of Arrival
UE	User Equipment

LIST OF FIGURES

Figure 2.1. Signal propagation in wireless medium 4

Figure 2.2. Two-ray ground reflection model..... 8

Figure 2.3. The received power referenced to the transmitted power as a function of the transmitter-receiver distance according to the free-space model..... 9

Figure 2.4. Diffraction 10

Figure 2.5. The Hata model 12

Figure 2.6. The Walfisch-Ikegami propagation model..... 13

Figure 3.1. CGI + TA method..... 16

Figure 3.2. MS-BTS geometry 18

Figure 3.3. TOA data fusion 21

Figure 3.4. Location determination using TDOA..... 24

Figure 3.5. AOA data fusion..... 26

Figure 3.6. A-GPS architecture..... 28

Figure 4.1. Graphical representation of an artificial neuron..... 31

Figure 4.2. Common activation functions 31

Figure 4.3. Multilayer perceptron network 32

Figure 4.4. Back-propagation of error 34

Figure 4.5. Back-propagation learning architecture in MLP 34

Figure 4.6. A decision tree 37

Figure 5.1. EARBALE architecture 42

Figure 5.2. Field measurement process 43

Figure 5.3. Data flow diagram for dataset acquisition 44

Figure 5.4. Dataset examination process in Google Earth™ software 45

Figure 5.5. Effect of using the minimum split criterion on mean square error 46

Figure 5.6. Resulting DT for feature selection (for $Q'=110$) 47

Figure 5.7. Three-layer ANN model used in EARBALE 47

Figure 5.8. ANN Classification Performance 49

Figure 5.9. Location estimation via circle intersection using RSS values from serving and neighbor BTSs 50

Figure 5.10. Singular BTS algorithm used as fall-back 51

Figure 6.1. Subscribers' TA distribution in real GSM network and in the dataset 53

Figure 6.2. Performance evaluation results 54

1. INTRODUCTION

Location Based Services enable personalized services to mobile subscribers based on their current position. They provide new opportunities for cellular operators as well as application and content providers for the provision of innovative value added services and creation of new revenue sources. Consequently, mobile positioning in wireless systems has received significant attention in both research and industry over the past few years since it plays a key role in providing LBS such as wireless emergency services, intelligent transportation systems, and location-based billing [1, 2].

Mobile positioning involves a variety of technologies, which are divided into two major categories: network-based and handset-based location estimation. Handset-based positioning methods require a modified handset to calculate its own position, for instance by using a GPS receiver embedded in the device [3]. The drawbacks of using handset-based methods are the cost of deploying new handsets, delay for the adoption of LBS due to slow spread of these new handsets and cost of developing a suitable low-power and economical integrated technology for the wireless communication systems. On the contrary, using network-based methods for mobile positioning in wireless communication systems has its advantages. When compared to the handset-based methods, the network-based methods are relatively less complex. However, although they can be used in many situations where GPS-based methods cannot be applied (such as indoor positioning), the mobile positioning in the network-based methods is generally less accurate than its counterpart in handset-based methods. Thus, the improvement of the accuracy for mobile positioning becomes an important issue especially for network-based solutions.

So far, a wide variety of localization techniques have been proposed using measurements performed within cellular networks, such as RSS, TOA, TDOA, AOA methods, and using the satellite-based GPS technology, either as stand-alone GPS or A-GPS [4, 5, 6, 7, 8].

The TOA method is able to locate legacy handsets but requires installation of new network elements called LMUs at each BTS. The AOA method requires antenna arrays at BTSs or terminals [9, 10]. In most systems, these arrays are usually arranged at BTSs because, for reasons of economy and complexity, a terminal-based solution is not practical. The A-GPS method requires installation of reference GPS receivers beside the integration of a GPS receiver into handsets.

An RSS-based method basically estimates handset's coordinates by fusing CGI, TA, and Rx level information. TA value corresponds to the time it takes for a signal to be transmitted from the MS to the serving BTS and it is used for synchronization purposes in GSM. Rx levels are measurements of the strength (i.e., power) of signals received by the MS from the serving BTS and from up to the six strongest neighbor BTSs in GSM. The level of a signal received by an MS, or more precisely the attenuation the signal has experienced, depends on the reciprocal position of the MS and the BTS from which the signal was transmitted. Attenuation values from multiple neighbor BTSs are modeled by the CI + TA + Rx level locationing method through basic prediction models (Okumura-Hata, Longley-Rice, Walfisch-Bertoni, etc.) and used to estimate the location of the MS.

Although location estimation algorithms based on signal attenuation may not be the most promising approach for providing location services, signal strength is the only common information available among various kinds of mobile network. Together with the fact that the geographic conditions and the cell layout in metropolitan areas are not the same, new approaches for location estimation algorithms based on signal attenuation have to be investigated [11].

In this study, we introduce a pattern recognition based environment-aware location estimation method, namely Environment Aware RSS Based Location Estimation (EARBALE), and evaluate its performance. We utilize signal measurements in a 900 MHz GSM network and data compiled from Istanbul. EARBALE method uses preprocessing and dimensionality reduction via DT on these bulk empirical data and apply artificial ANN based pattern recognition for the adoption of the most appropriate wireless channel model. By this way, our trained ANN is capable of identifying the environment of an input measurement as urban, suburban, or rural. An urban area is an area with an increased

density of human-created structures in comparison to the areas surrounding it. A suburban area is an inhabited district located either inside a town or city's limits or just outside of it. A rural area is a sparsely settled place away from the influence of large cities. According to the identification of the most probable environment, the localization algorithm uses the corresponding Hata propagation model [12] in the triangulation phase. We evaluate and compare EARBALE to the generic RSS based localization method that uses triangulation with Hata's default propagation model. By this way, we investigate the importance of environment estimation for any RSS based localization algorithm.

This study is organized as follows: In the following two chapters, two important background topics related to the thesis study are presented: Chapter 2 presents the principles of radio wave propagation and in Chapter 3 the most important location estimation techniques existing in literature are presented. Chapter 4 describes the two machine learning techniques used as the backbone of EARBALE, namely ANNs and DTs. In Chapter 5, the proposed environment-aware algorithm together with its general architecture, phases, and key functions is described. Performance evaluation of the proposed algorithm is presented in Chapter 6 with illustrations and comments on them. And finally with Chapter 7, the study is finalized with concluding remarks.

2. RADIO WAVE PROPAGATION

With location estimation systems based on radio signals, it is important to know the propagation properties of electromagnetic radiation. Phenomena such as signal attenuation, reflection, scattering and diffraction have important roles in location estimation. These are depicted in Figure 2.1. Their importance is emphasized in non-satellite systems which have to operate in complex propagation environments, such as urban or mountainous areas. This chapter addresses the most important theoretical aspects of radio wave propagation and reviews some propagation models based on them.

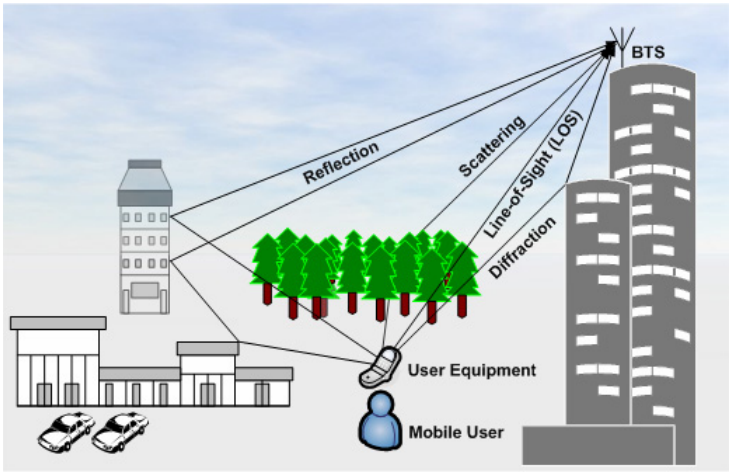


Figure 2.1. Signal propagation in wireless medium

2.1. Basic Principles

The basic concept in the theory of electromagnetic radiation is an *electric field*, which is always related to electric current [13]. An electric field E is defined by its direction and magnitude at each point. The magnitude, denoted by $|E|$ is measured in units of volts per meter (V/m). Periodic fluctuations of an electric field are called *radio waves*. Radio waves can be decomposed into orthogonal components, typically the horizontal and the vertical component. The ratio of the magnitudes of the two components, or equally the direction of the electric field defines the *polarization* of the wave. For

instance, if the magnitude of the vertical component is always zero, i.e. the direction vector is always parallel to the horizontal axis, the wave is said to be horizontally polarized.

An electric field corresponds to a *power density flow F*, measured in watts per square meter (W/m^2), which is proportional to the square of the magnitude of the electric field [13]. Given the power density flow, the gain of a receiving antenna, G_r which depends on the physical size of the antenna and frequency, the wave length λ, and the system hardware loss, L, the received power is given by

$$P_R = \frac{FG_r\lambda^2}{4\pi L} \tag{2.1}$$

Even though the wave length λ appears in Equation (2.1), it does not follow that the received power would increase proportionally to the square of the wave length, because the wave length also affects the gain of the receiving antenna G_r. In fact, if the physical size of the antenna and the power density flow are constant, the wave length terms cancel each other out, and thus the received power is independent of the frequency. However, the frequency can indeed affect the power density flow due to interactions with the propagation medium.

Because the values of received power vary over a wide range, it is convenient to use logarithmic scale. A ratio of two quantities can be presented in decibels (dB) which indicates the logarithm of the ratio multiplied by ten. The unit of decibelwatt (dBW) is the ratio of power referenced to one watt. Conversions between watts and decibelwatts are made with the following two equations:

$$P[\mathrm{dBW}] = 10\log\left(P[\mathrm{W}]\right), \tag{2.2}$$

$$P[\mathrm{W}] = 10^{\frac{P[\mathrm{dBW}]}{10}} \tag{2.3}$$

For instance, 0 dBW is equal to one watt, 10 dBW is equal to 10 watts, 20 dBW is equal to 100 watts, etc. The unit of decibelmilliwatt (dBm) is defined similarly as the ratio

of power referenced to one milliwatt. Conversions between two decibel units, for instance, decibelwatts and decibelmilliwatts, can always be performed simply by adding a constant to the original value. The following two equations are used for converting decibelwatts to decibelmilliwatts and vice versa:

$$P[\text{dBm}] = P[\text{dBW}] + 30, \tag{2.4}$$

$$P[\text{dBW}] = P[\text{dBm}] - 30. \tag{2.5}$$

2.1.1. Free-space Attenuation

Because a wave front proceeds in three dimensions, the maximum received power at distance d must decrease in the inverse of the area of a sphere with radius d. If the absorption loss of the propagation medium is ignored, the power density flow, F, is given by

$$F = \frac{P_T G_T}{4\pi d^2}, \tag{2.6}$$

where P_T is the transmitted power, G_T is a factor depending on the transmitting antenna, and d is the distance. Combining Equations (2.1) and (2.6) gives the received power, which is usually given in decibels:

$$P_R[\text{dB}] = P_T[\text{dB}] + 10\log(G_T) + 10\log(G_R) + 20\log(\lambda) - 20\log(d) - 22.0. \tag{2.7}$$

Equations (2.6) and (2.7) are valid only in free-space environment, where there are neither reflections, absorption, diffraction nor other distortions. If the LOS between the transmitter and the receiver is obstructed, the received signal power is significantly lower than the free-space equations suggest. Furthermore, they do not necessarily give a good approximation even in LOS conditions [13].

2.1.2. Absorption

In any real-world communication system, the signals propagate in some medium. In wireless terrestrial systems, that medium is mainly the atmosphere and, in lesser degree, materials such as glass, concrete, wood, etc. Due to interactions with the medium, the signal loses a certain proportion of its remaining energy on every unit of distance it propagates. Thus, absorption causes the power density flow to decrease proportionally to γ^{-d}, where d is the distance, and γ is a constant depending on the properties of the medium and signal frequency. This means that in decibel scale, the loss is linear with respect to the distance.

Absorption loss is particularly great in the upper microwave region, where the frequencies are above 10 GHz. With these frequencies the absorption due to atmosphere becomes comparable to the free-space attenuation, especially in heavy rain conditions and with long transmitter-receiver distances. With frequencies used in most wireless communication systems, below 10 GHz, the atmospheric absorption is insignificant with distances up to 10 km.

Absorption caused by other media than air is generally very strong. Moreover, in addition to absorption, obstructions cause the wave to be reflected, which further decreases the amount of energy passing through. Taking into account both reflection and absorption, the total attenuation per obstruction is typically 1–20 dB below 10 GHz, and 1–60 dB above 10 GHz [13].

2.1.3. Reflection

Reflection occurs when a wave meets an obstacle with size much bigger than the wave length. The part of the wave that is not reflected back loses some of its energy by absorbing to the material and the remaining part passes through the reflecting object. In terrestrial communication systems the waves usually reflect from ground, producing a two-ray path between the transmitter and the receiver, shown in Figure 2.2. The *plane of incidence* is defined as the plane containing both the incident ray and the reflected ray, and the *angle of incidence* is the angle between the reflecting surface and the incident ray.

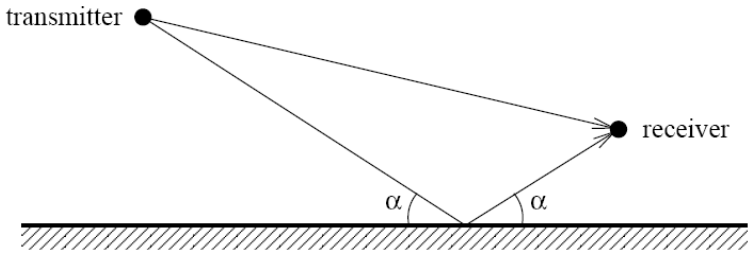


Figure 2.2. Two-ray ground reflection model

The received signal consists of the direct LOS ray and the reflected ray. The two rays arriving to a receiver can have different phase and in the worst case they cancel each other out. The magnitude of the reflected signal depends on the *Fresnel reflection coefficient* [14], which depends on the properties of the reflecting ground, the frequency of the wave, and the angle of incidence. Roughness of the reflecting surface causes the propagating waves to *scatter* in all directions, and therefore, the reflection coefficient of a rough surface is smaller than its identical but flat surface. In general, the reflection coefficient is different for the vertical and the horizontal component of the wave. In such cases, reflection can change wave polarization [13].

Figure 2.3. presents the attenuation curve of the two-ray model with certain parameters. It can be seen from the figure that with long distances the two-ray model coincides with the *fourth-power approximation*, which is given by

$$P_R[\mathrm{dB}] = P_T[\mathrm{dB}] + 10\log(G_T) + 10\log(G_R) - 40\log(d) - 22.0, \qquad (2.8)$$

where the received power is proportional to the inverse of the *fourth* power of the distance rather than the square of the distance which appears in the free-space model.

The parameters used in Figure 2.3 are: transmitter elevation is 50 meters, receiver elevation is two meters, frequency is 900 MHz, relative permittivity of the ground is 15, and antenna gains and system loss are 1.0 (no loss).

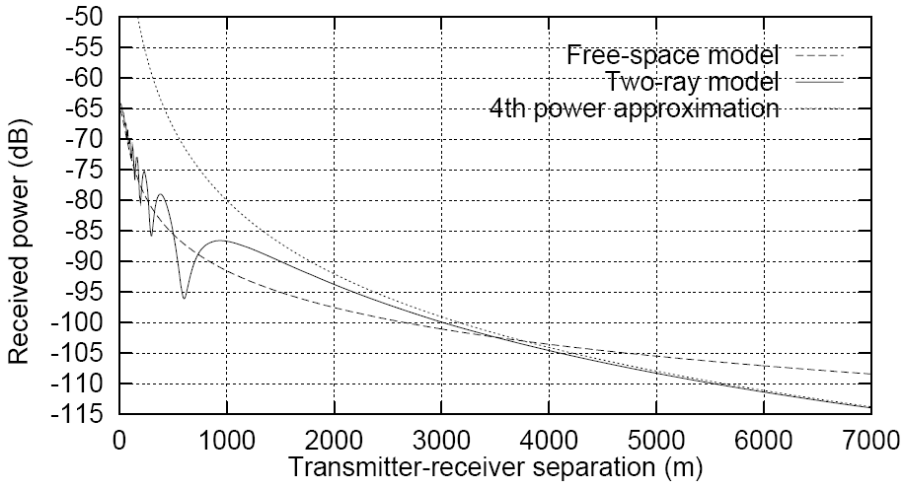


Figure 2.3. The received power referenced to the transmitted power as a function of the transmitter-receiver distance according to the free-space model (Equation (2.7)), the two-ray model, and the fourth-power approximation of the two-ray model (Equation (2.8)).

2.1.4. Diffraction

According to Huygen's principle [14], all points on a wavefront are point sources of secondary waves propagating to all directions. Therefore, each time a radio wave passes an edge such as a corner of a building the wave bends around the edge and continues to propagate into the area shadowed by the edge. This effect is called *diffraction*. In Figure 2.4, the transmitter is situated near an obstacle. The arrows describing the direction of propagation indicate how the signal reaches the areas around the corner due to a source of secondary waves situated at the corner of the obstacle. Note that the single source of secondary waves shown in Figure 2.4 is only one of the infinite number of such sources on the wavefront.

The more the waves have to bend around a corner, the more they lose their energy. Therefore the areas to which the rays have to bend more, gain relatively less additional field strength than the areas to which the rays can proceed almost linearly. The field strength of the secondary waves is much smaller than the one of the primary waves. In practice, the diffracted waves can be neglected if there is a LOS between the transmitter and the receiver.

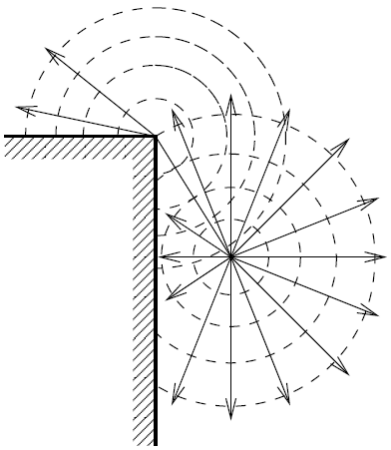


Figure 2.4. Diffraction

2.2. Propagation Models

Prediction of radio wave propagation is useful in activities such as allocation of bandwidth, cell planning and location estimation. Propagation models are used to predict the properties of the propagating waves, usually the received signal power and its variability. It is also possible to predict polarization, time dispersion, frequency selectivity and other properties that affect the performance of communication systems [13].

General propagation models typically describe the field strength as a function of the distance between the transmitter and the receiver. The two-ray ground reflection model and its approximations described in the previous section are examples of general models. The models usually include some factors such as ground properties, or the Fresnel reflection coefficient in the two ray model. One of the most popular general models is the Okumura model. It consists of the free space loss (Equation 2.7) and a correction factor. The correction factor is given by a function of the distance and the frequency of the signal. Okumura presented the functions graphically as curves. Different curves exist for open, quasi-open, suburban and urban area. Later, Hata used the empirical data presented by Okumura and gave mathematical formulae known as the Hata model, which closely fits Okumura's curves. The Hata model was later extended to higher frequencies in order to include the 1800 MHz frequency band used in some cellular telephone systems.

Various wireless radio propagation models exist in the literature [12, 15]. These models are basically divided into two distinct classes, namely deterministic and stochastic. The deterministic model corresponds to a deterministic description of the environment that causes path loss and multipath effects, and it is very useful when multipath is caused by a small number of paths that can be accurately characterized. The stochastic model corresponds to a statistical model of the environment and it is most appropriate when the multipath effects are caused by a large number of paths between transmitter and receiver that would be impossible to characterize in a deterministic way [16].

2.2.1. Hata Model

Hata model is the most popular channel model and best suited for a large cell coverage (distances up to 100 km) and it can extrapolate predictions up to the 2GHz band [12]. This model has been proven to be accurate and is used by many computer simulation tools.

There are three adoptions of the model to three main environments, which are urban, suburban, and rural areas. Our DT assisted ANN architecture selects one of these models in accordance with the given input data, which will be explained in Chapter 5 in detail. The analytical approaches for the three versions of the Hata model are listed below. Formulae (2.9) and (2.10) are for urban areas, formula (2.11) is for suburban areas, and formula (2.12) is for rural areas.

$$PL_{urban} = 69.55 + 26.16\log(f) - 13.82\log(h_t) - a(h_m) + [44.9 - 6.55\log(h_t)]\log(d) \quad (2.9)$$

$$a(h_m) = 3.2\log^2(11.75 \times h_m) - 4.97 \quad (2.10)$$

$$PL_{suburban} = PL_{urban} - 2\log^2(f/28) - 5.4 \quad (2.11)$$

$$PL_{open\ area} = PL_{urban} - 4.78\log^2(f) + 18.33\log(f) - 40.94 \quad (2.12)$$

for a large city where

- PL – path loss (dB)
- f – operating frequency (MHz)
- $a(h_m)$ – correction factor for mobile unit antenna height (dB)
- d – distance from transmitting station (m)
- h_t – transmitting station antenna height (m)
- h_m – mobile unit antenna height (m)

Using the following parameters: f = 2000 MHz, h_t = 40 m, h_m = 1.5 m, the loss predictions for the urban Hata model is ilustrated graphically in Figure 2.5.

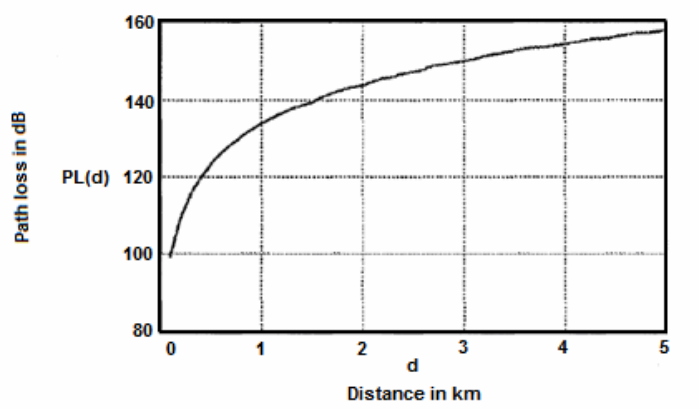


Figure 2.5. The Hata model

In the proposed EARBALE scheme, we utilize deterministic Hata models and use them to solve the unknown distance variable from the received signal amplitude together with some GSM network parameters such as TA and CGI.

2.2.2. Walfisch-Bertoni Model

Another popular model is Walfisch-Bertoni model [15]. In this model for path loss prediction, over a frequency range of 800 to 2000 MHz, a distinction is made between LOS and NLOS situations. It is the first theoretical model to explain and address the

effects of buildings on propagation in urban areas in the 300 MHz - 30 GHz band. The model contains three equations: free space loss, roof-to-street diffraction and scatter loss, and multiscreen loss. This model is depicted in Figure 2.6. and its analytical approach is as follows:

$$L_f = 32.4 \ + 20\log(d) + 20\log(f)dB\,, \tag{2.13}$$

$$L_{rts} = (-16.9) - 10\log(W) + 10\log(f) + 20\log(\Delta h_m) + L_0 dB\,, \tag{2.14}$$

$$L_{ms} = L_{bsh} + k_a + k_d \log(d) + k_f \log(f) - 9\log(b)\,, \tag{2.15}$$

$$PL_1 = L_f + L_{rts} + L_{ms}\,. \tag{2.16}$$

where parameters are shown in Figure 2.6.

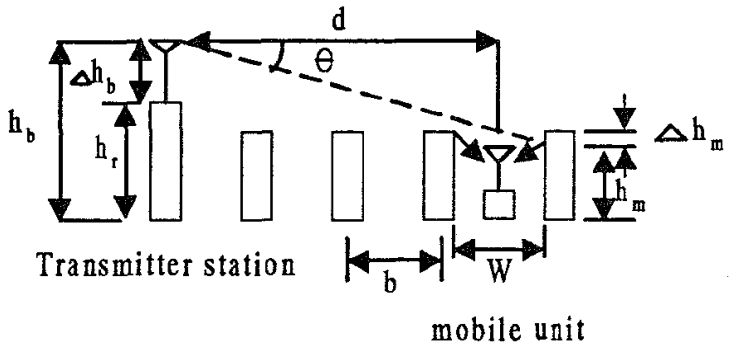


Figure 2.6. The Walfisch-Bertoni propagation model

The Walfisch-Bertoni model shows how the buildings influence the propagation and hence identifies those physical properties that are significant [15]. However, it is suitable for only urban areas where a detailed knowledge such as street widths, building heights, and so on exists. We do not use this model since we do not have a database of

street and building details for our empirical dataset. In addition, most of the urban cities have areas that possess suburban or even rural characteristics, making the use of Walfisch-Bertoni model problematic.

3. MOBILE POSITIONING TECHNIQUES

Since cellular networks have not originally been designed to support mobile positioning, there are many localization techniques based on either the installation of new equipments to the network such as LMUs and antenna arrays or making use of some available parameters such as CGI and TA. Methods requring the installation of new equipments to the network such as AOA or A-GPS provide better localization accuracy at the cost of money. Before presenting the details of mobile positioning techniques in literature, we describe the categories that these techniques are divided into. Mainly, the localization techniques are categorized as network-based, mobile-based, or mobile-assisted.

In network-based positioning methods, one or more BTSs make the required measurements and send these measurement results to a locationing center where the position of the MS is estimated. The most important advantage of network-based methods is that they do not require any changes to existing handsets. However, for network-based locationing systems to work properly, the MS in consideration must be in active mode to enable location measurements. Positioning in idle mode is not possible.

In mobile-based positioning methods, unlike the previous case the MS performs the measurements and makes the location estimation by itself. In this method, some extra information such as BTS coordinates might be needed from the network for the mobile unit to be able to estimate its location. Mobile-based methods are the most accurate among other methods but they do not operate on legacy handsets and hence require new mobile terminals supporting these features.

In mobile-assisted positioning, the MS makes the measurements as in the mobile-based positioning but sends these measurement results to a locationing center for further processing. By this way, computational weight on the handset is diminished and better accuracy is achieved.

Another important categorization for location estimation techniques is based on the measurement principle. According to this, the measurement criteria of any localization method belongs to multilateral, unilateral or bilateral category. In multilateral techniques, which corresponds to the network-based localization methods, several BTSs make measurements together. In unilateral techniques, which correspond to mobile-based or mobile-assisted localization methods, MS takes measurements from several BTSs. Finally, in bilateral techniques either MS takes measurements from a single BTS or a single BTS takes measurements from an MS.

3.1. Cell Global Identification and Timing Advance

Using CGI together with TA parameter for location estimation is one of the simplest methods in wireless location estimation because all the required parameters for its operation reside in the network and it does not require any changes to the hardware or software system in the network.

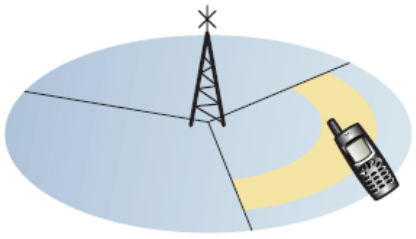


Figure 3.1. CGI + TA method

CGI identifies the cell the MS is located in. A cell can be a circular (omni-directional) or a triangular sector as shown in Figure 3.1. The TA parameter is an estimate of the distance from the MS to the serving BTS and it is a parameter used to avoid overlapping of bursts at BTS. It is proportional to the distance between the MS and BTS. TA values are divided into 64 slots (0-63), each with a radius of 550 meters. By using the TA value, the location of the MS can be constrained further than the CGI as it can be narrowed to a circle (in case of an omni-directional cell) or a sector (in case of a sectoral cell) in steps of a 550 meters radius from the BTS.

The accuracy of this method depends on the cell size, which may vary from 100 meters in urban areas to 35 kilometers in rural areas. In cells that cover a limited geographical area, the accuracy is fairly good, but it decreases rapidly as the distance between the transmitter and receiver increases. Accuracy will also depend on whether the cell is an omni-cell or a triangular sector cell.

This is a bilateral locationing method that can be implemented as a network-based or mobile-based scheme. In addition to its simplicity, another advantage is that no calculations are needed to obtain location estimate. Thus, CGI + TA based locationing is fast and suitable for applications requiring high capacity. Its major drawback is that the accuracy directly depends on the cell radius. In dense urban areas location accuracy is better due to small cell sizes.

3.2. Signal Strength

Using signal strength measurements from the control channels of several BTSs, the distances between the MS and the BTSs can be estimated [17]. Assuming two-dimensional geometry, an omnidirectional BTS antenna, and free-space propagation conditions, signal level contours around BTSs are circles. If signal levels from three different BTSs are known, the location of the MS can be determined as the unique intersection point of the three circles. However, practical propagation conditions especially in urban areas are far from free-space propagation. Therefore, an environment-dependent propagation model for the dependence of received signal level on BTS – MS distance should be used. In urban areas, the received signal level generally decreases more rapidly with distance than in open areas.

Multipath fading and shadowing pose a challenge for distance estimation based on signal level. The instantaneous, narrowband signal level may vary by as much as 30-40 dB over a distance of only a fraction of the wavelength. Random variations of this order of magnitude cause very large errors in distance estimates. However, fast fading can be smoothed out by averaging the signal strength over time and frequency band. Time-averaging only has a minor effect, due to the motion in the surrounding environment, if the MS is stationary. Contrary to fast fading, the random variations caused by shadowing can

not be compensated. Thus, the variations in antenna orientation and local shadowing conditions around the MS (indoors, inside a vehicle etc.) are seen as random errors in distance estimates and consequently in position estimate. Location accuracy also depends on the accuracy of the propagation model and the number of available measurements.

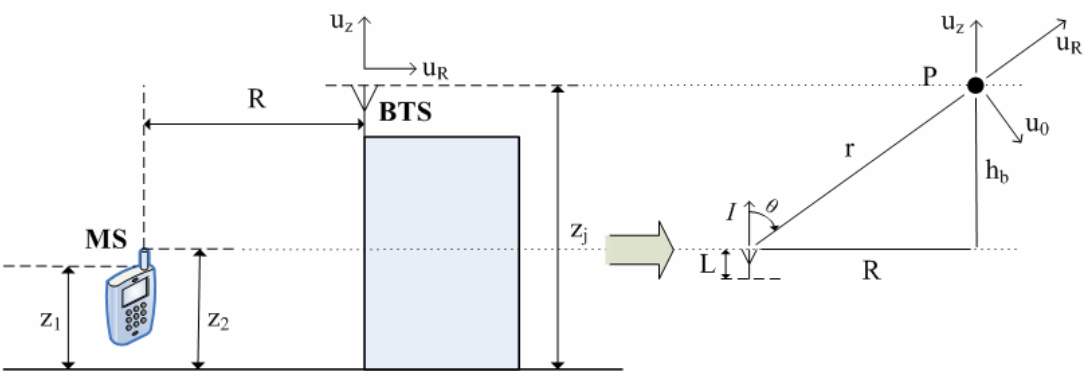


Figure 3.2. MS-BTS geometry

The geometry of the MS and BTS is given in Figure 3.2. The electric field [18] radiated by the antenna of the MS can be written as:

$$E(r,\theta) = j\frac{60I_0}{r}\left[\frac{\cos\left(\frac{kL}{2}\cos\theta\right)-\cos\left(\frac{kL}{2}\right)}{\sin\theta}\right]e^{-jkr} \tag{3.1}$$

The current can be expressed by I_0 cos(wt-kr) where I_0 is the amplitude and w is the frequency. Therefore, the electric field strength can be expressed in the following form

$$\left|E(r,\theta)\right| = \frac{E_0}{r}\left[\frac{\cos\left(\frac{kL}{2}\cos\theta\right)-\cos\left(\frac{kL}{2}\right)}{\sin\theta}\right] \tag{3.2}$$

$$\left|E(R)\right| = E_0 \left[\frac{\cos\left(\frac{kL}{2}\frac{h}{\sqrt{R^2+h^2}}\right) - \cos\left(\frac{kL}{2}\right)}{R} \right] \tag{3.3}$$

where

$$E_0 = 60 I_0 ,\ I_0 = \sqrt{\frac{P}{Rs}} \tag{3.4}$$

Here, *P* is the signal power radiated by the MS and *Rs* is the radiation resistance of the MS antenna [18].

For the locationing process using signal strength values to work, a minimum of three BTSs are required in order to determine the two dimensional position of the MS by using triangulation. The first step is to get the three nearby BTSs from their received field strengths. The theoretical and measured electric field strengths received at BTS for MS are

$$E = \left[E_1, E_2, E_3\right] \tag{3.5}$$

$$E_m = \left[E_{m1}, E_{m2}, E_{m3}\right] = \alpha E + n \tag{3.6}$$

respectively, where α stands for the signal fading, n for the channel noise.

The next step is to obtain the distance between the MS and BTS. This calculation is done by using ML estimation method.

$$R = \left[R_1, R_2, R_3\right] \tag{3.7}$$

The least square method is then used to obtain the position of the MS from the estimating distance. The electric field strength *E(R)* at a distance *r* from the MS is given by

$$E(R) = E(R_0) + \frac{\partial E}{\partial R}\Delta R + O(\Delta R) = E_0 + J_0 \Delta R + O(\Delta R) \tag{3.8}$$

where $O(\Delta R)$ can be ignored when ΔR is very small when compared to R and

$$J_0 = \begin{bmatrix} \frac{\partial E_1}{\partial R_1} & & 0 \\ & \frac{\partial E_2}{\partial R_2} & \\ 0 & & \frac{\partial E_3}{\partial R_3} \end{bmatrix} \tag{3.9}$$

From (3.3) we have

$$\frac{\partial E}{\partial R} = E_0 \left[\frac{kLh}{2\left(R^2 + h^2\right)^{3/2}} \sin\left(\frac{kLh}{2\sqrt{R^2 + h^2}} \right) - \frac{1}{R^2} \left(\cos\left(\frac{kLh}{2\sqrt{R^2 + h^2}} \right) - \cos\left(\frac{kL}{2} \right) \right) \right] \tag{3.10}$$

The error between measurement and theoretical value for the electric field strength is given as

$$\varepsilon = E_m - E\left(R\right) = E_m - E_0 - J_0 \Delta R \tag{3.11}$$

The minimum of ε^2 is obtained when

$$\frac{\partial u}{\partial R} = 0 \tag{3.12}$$

The main iteration equation for R is obtained as

$$R^{k+1} = R^k + \left(J_k^T J_k\right)^{-1} J_k^T \left(E_m - E^k\right) \tag{3.13}$$

Then, the position of the MS is given by the intersection of the circles [18].

Signal strength method is unilateral and can be implemented as mobile-assisted or mobile-based method. Mobile-based implementation requires that BTS coordinates are transmitted to the MS. Signal strength method is easy to implement in GSM, based on measurement reports that are continuously transmitted from the MS back to the network in active mode. Therefore, it does not require any changes to existing handsets, and is often called a network-based method although it is the MS that performs the measurements. An alternative implementation is to modify MSs to enable sending measurement reports in idle mode also. GSM phones with this capability are already available. Signal strength is an easy and low-cost method to enhance the accuracy of CGI + TA based locationing described in the previous section [17].

3.3. Time of Arrival

Time of arrival technique uses the one way trip time of a burst signal between a BTS and an MS. In this method, each BTS waits for the burst signal sent at a certain time by the MS and records the arrival times. After that, these BTSs send their data to a locationing center for locationing calculations. Similarly, in some TOA architectures the MS gets the bursts from BTSs and sends these data to the locationing center. Since the measurements may both be made by the MS or the BTSs, TOA method can both be categorized as a unilateral mobile-assisted technique or a multilateral network-based technique, respectively.

With three BTSs, as shown in Figure 3.3, the coordinates (x_m, y_m) of an MS can be calculated using the least squares method [19]. The locationing center makes the calculations and estimates the coordinates of the MS using triangulation as described below.

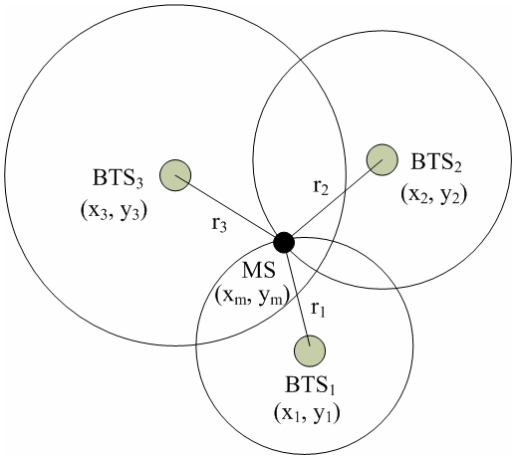


Figure 3.3. TOA data fusion

The distance r_i between the MS and BTS_i can be simply calculated as

$$r_i = \left(t_i - t^0\right)c \tag{3.14}$$

where c is the speed of light, t^0 is the time instant that the burst signal is transmitted, and t_i is the TOA of the received signal. The coordinates (x_m, y_m) of the mobile user can be found by making use of the following distance equations:

$$r_1^2 = x_m^2 + y_m^2 \tag{3.15}$$

$$r_2^2 = \left(x_2 - x_m\right)^2 + \left(y_2 - y_m\right)^2 \tag{3.16}$$

$$r_3^2 = \left(x_3 - x_m\right)^2 + \left(y_3 - y_m\right)^2 \tag{3.17}$$

Subtracting (3.15) from (3.16) gives

$$r_2^2 - r_1^2 = x_2^2 - 2x_2x_m + y_2^2 - 2y_2y_m \tag{3.18}$$

Similarly, subtracting (3.15) from (3.17) gives

$$r_3^2 - r_1^2 = x_3^2 - 2x_3x_m + y_3^2 - 2y_3y_m \tag{3.19}$$

The above two equations can be written in matrix form as

$$\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \end{bmatrix} = \frac{1}{2} \begin{bmatrix} K_2^2 - r_2^2 + r_1^2 \\ K_3^2 - r_3^2 + r_1^2 \end{bmatrix} \tag{3.20}$$

where

$$K_i^2 = x_i^2 + y_i^2 \tag{3.21}$$

Then, (3.20) can be rewritten as

$$\mathbf{Hx = b} \tag{3.22}$$

where

$$\mathbf{H} = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} K_2^2 - r_2^2 + r_1^2 \\ K_3^2 - r_3^2 + r_1^2 \end{bmatrix} \tag{3.23}$$

Then, the least-squares solution is

$$\hat{\mathrm{x}} = \left(\mathrm{H}^{\mathrm{T}}\mathrm{H}\right)^{-1} \mathrm{H}^{\mathrm{T}}\mathrm{b} \tag{3.24}$$

For this method to work, BTSs and the MS must be synchronized. Even very tiny errors in this synchronization will yield very large errors in the estimation process. Most of the time, the MS cannot be synchronized precisely enough to directly obtain the propagation time of the burst signal to travel between itself and a BTS. One solution to avoid this synchronization problem is to use differences in the time delays of several BTSs

instead of exact times. Time differences are used in the TDOA method which will be discussed in the next section.

3.4. Time Difference of Arrival

Rather than absolute times, TDOA method uses time difference values in the signal propagation time between the MS and BTSs. The main goal of the technique is to eliminate the synchronization problem with the MS. TDOA values of two BTSs form hyperbolae as shown in Figure 3.4.

In this figure, hyperbolic curves r_{ij} are defined as:

$$r_{ij} = \sqrt{\left(x_i - x_m\right)^2 + \left(y_i - y_m\right)^2} - \sqrt{\left(x_j - x_m\right)^2 + \left(y_j - y_m\right)^2} \tag{3.25}$$

where r_{ij} is the difference in the time delays between the MS and BTSs i and j.

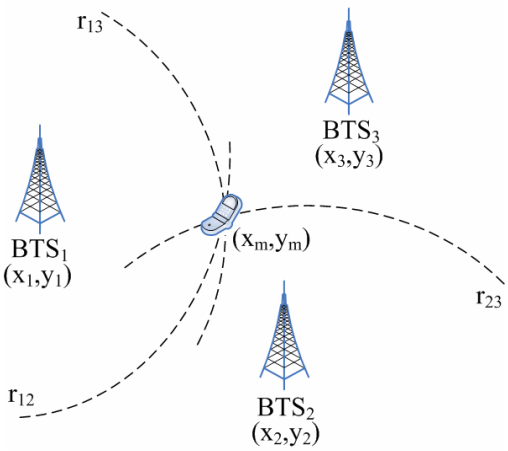


Figure 3.4. Location determination using TDOA

The time difference between the TOAs of the MS signal at BS_i and BS_1 is [19]

$$t_i - t_1 \tag{3.26}$$

And the distance differences are

$$r_{i1} = r_i - r_1 = \left(t_i - t^0\right)c - \left(t_1 - t^0\right)c = \left(t_i - t_1\right)c \tag{3.27}$$

Here we see that as we take the time differences, the t^0 values cancel each other, eliminating the need for MS synchronization. Rewriting the equation (3.16) in terms of TDOA measurement r_{21}:

$$\left(r_{21} + r_1\right)^2 = K_2^2 - 2x_2x_m - 2y_2y_m + r_1^2 \tag{3.28}$$

Rearranging:

$$-x_2x_m - y_2y_m = r_{21}r_1 + \frac{1}{2}\left(r_{21}^2 - K_2^2\right) \tag{3.29}$$

Rewriting the equation (3.17) in terms of TDOA measurement r_{31}:

$$-x_3x_m - y_3y_m = r_{31}r_1 + \frac{1}{2}\left(r_{31}^2 - K_3^2\right) \tag{3.30}$$

Rewriting the equations (3.29) and (3.30) in the matrix form:

$$\mathbf{Hx} = r_1\mathbf{c} + \mathbf{d} \tag{3.31}$$

where

$$\mathbf{c} = \begin{bmatrix} -r_{21} \\ -r_{31} \end{bmatrix},\ \mathbf{d} = \frac{1}{2}\begin{bmatrix} K_2^2 - K_{21}^2 \\ K_3^2 - K_{31}^2 \end{bmatrix} \tag{3.32}$$

After that, equation (3.31) can be used to solve for x:

$$\mathbf{x} = r_1\mathbf{H}^{-1}\mathbf{c} + \mathbf{H}^{-1}\mathbf{d} \tag{3.33}$$

Final solution for x is acquired by finding the r_1 value by substituting the equation (3.33) into equation (3.15). The equation (3.31) still holds in the existence of more than three BTSs with:

$$\mathbf{H}=\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \\ \cdots & \end{bmatrix},\ \mathbf{c}=\begin{bmatrix} -r_{21} \\ -r_{31} \\ -r_{41} \\ \cdots \end{bmatrix},\ \mathbf{d}=\frac{1}{2}\begin{bmatrix} K_2^2-K_{21}^2 \\ K_3^2-K_{31}^2 \\ K_4^2-K_{41}^2 \\ \cdots \end{bmatrix} \tag{3.34}$$

similarly yielding:

$$\hat{\mathbf{x}}=\left(\mathbf{H}^{\mathbf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathbf{T}}\left(r_1\mathbf{c}+\mathbf{d}\right) \tag{3.35}$$

3.5. Angle of Arrival

Angle of arrival method estimates the location of an MS by fusing the signal angle of arrival information from two BTSs. Angle measurements require additional hardware, such as antenna arrays to be installed to the network. It is a multilateral and network-based technique as several BTSs take measurements coming from the MS.

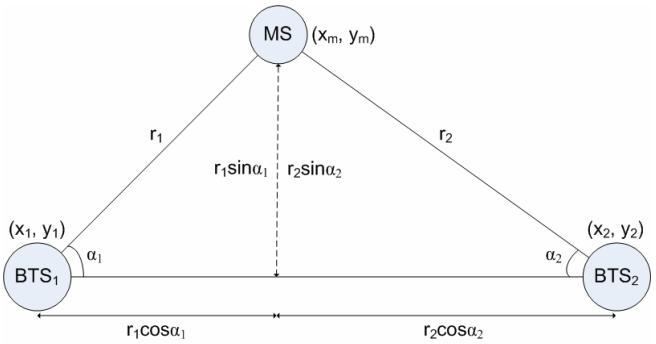


Figure 3.5. AOA data fusion

By combining the knowledge of the AOA from two BTSs as shown in Figure 3.5, the location of the MS can be calculated as follows [19]:

From Figure 3.5, we can write

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} r_1 \cos\alpha_1 \\ r_1 \sin\alpha_1 \end{bmatrix} \tag{3.36}$$

and

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} r_2 \cos\alpha_2 \\ r_2 \sin\alpha_2 \end{bmatrix} \tag{3.37}$$

Similarly, for any other BTS we can write

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} r_i \cos\alpha_i \\ r_i \sin\alpha_i \end{bmatrix} \tag{3.38}$$

These equations can be written as

$$\mathbf{Hx} = \mathbf{b}\,, \tag{3.39}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 0 & 1 \\ \hline \cdots \\ \hline 1 & 0 \\ 0 & 1 \end{bmatrix},\ \mathbf{x} = \begin{bmatrix} x_m \\ y_m \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} r_1 \cos\alpha_1 \\ r_1 \sin\alpha_1 \\ \hline x_2 + r_2 \cos\alpha_2 \\ y_2 + r_2 \sin\alpha_2 \\ \hline \cdots \\ \hline x_n + r_n \cos\alpha_n \\ y_n + r_n \sin\alpha_n \end{bmatrix} \tag{3.40}$$

Then the least squares solution for x is

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{b} \tag{3.41}$$

AOA works best in LOS conditions and its accuracy is highly affected in conditions where there is no LOS signal between MS and BTSs. In dense urban areas, there are generally no LOS signal because of the obstacles in the environment. Therefore, AOA method works better in rural and suburban areas. In addition, the angular orientation of the BTS antenna arrays must be calibrated very precisely because a small error in the angle may yield a large error in the localization of the MS.

3.6. Assisted GPS

A-GPS describes a system where outside sources, such as an assistance server and reference network, help a GPS receiver perform the tasks required to make range measurements and position solutions. The assistance server has the ability to access information from the reference network and also has computing power far beyond that of the GPS receiver. The assistance server communicates with the GPS receiver on the handset via a wireless link. With assistance from the network, the receiver can operate more quickly and efficiently than it would unassisted, because a set of tasks that it would normally handle is shared with the assistance server. The resulting A-GPS system, consisting of the integrated GPS receiver and network components, boosts performance beyond that of the same receiver in a stand-alone mode. Since the A-GPS receiver and the assistance server share tasks, the process is quicker and more efficient than regular GPS, but it is dependent on cellular coverage. Figure 3.6 shows the general architecture of an A-GPS system.

There are three basic types of data that the assistance server provides to the GPS receiver: precise GPS satellite orbit and clock information; initial position and time estimate; and for A-GPS-only receivers, satellite selection, range, and range-rate information. The assistance server is also able to compute position solutions, leaving the GPS receiver with the sole job of collecting range measurements [20].

Since GPS uses higher frequency levels than that of GSM, mobile handsets must have two antennae for both. In addition to this, GPS-enabled handsets must have higher battery capacities than legacy handsets because of the relatively high battery consumption of GPS receivers.

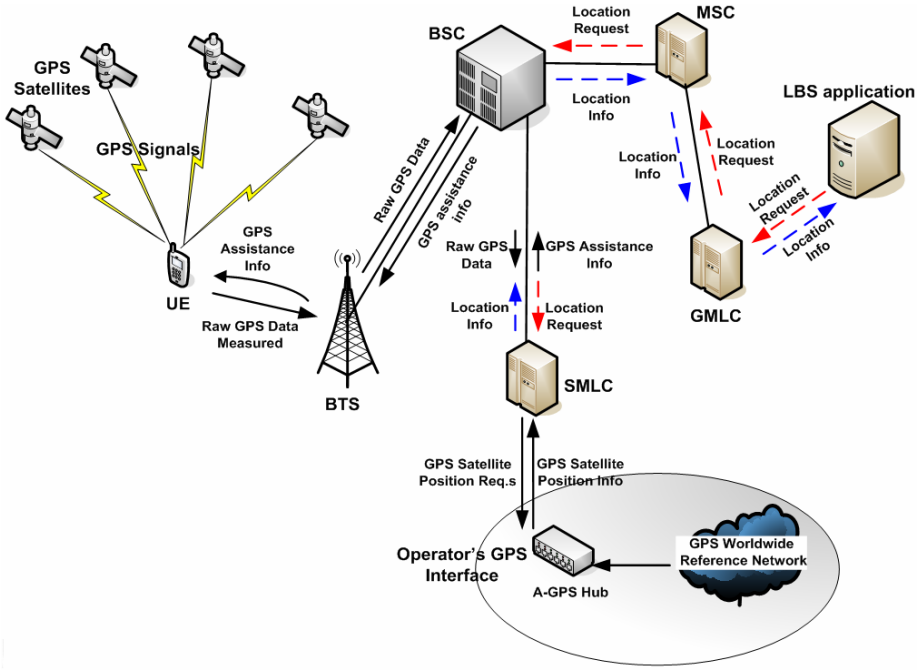


Figure 3.6. A-GPS architecture

3.7. Database Correlation

Database correlation method [17] is a generic location method that can be applied to any cellular network. The key idea is to store the signal information seen by an MS, from the whole coverage area of the location system, in a database that is used by a location server. The database should contain signal information samples, called *fingerprints*, with a resolution comparable to the accuracy that can be achieved with the method, and this resolution may vary in different environments. Depending on the particular cellular system, the signal fingerprints could include signal strength, signal time delay, or even channel impulse response. Any location-dependent signal information that can be measured by the MS is useful for the DCM technique. Also, it is possible to use measurements performed by the network as well as by the MS. When the MS needs to be located, the necessary measurements are performed and transmitted to the location server. The location server then calculates the MS location by comparing the transmitted fingerprint and the fingerprints of the database. This fingerprint may consist of signals measured from the network such as GSM or GPS. The location server must be powerful enough to process all location requests in a reasonable time. In a large-scale implementation, this may require distributed processing.

The major effort in applying DCM is the creation and maintenance of the database. The signal fingerprints for the database can be collected either by measurements or by a computational network planning tool. Measurements are harder and more time consuming but produce more accurate fingerprint data. Also a combination of measured and computed fingerprints can be used. The only assumption is that the database contains up-to-date data. However, minor changes in the network or propagation environment, e.g. new buildings, will only be seen as lowered location accuracy if the database is not updated. Also, it should be noted that similar information that is contained in the DCM database is also needed in network planning. Therefore, the creation and maintenance of the database also support network planning.

4. MACHINE LEARNING TECHNIQUES USED IN EARBALE

This chapter presents the two machine learning techniques used as the backbone of EARBALE, namely ANNs and DTs. As stated before, EARBALE uses preprocessing and dimensionality reduction via DT on empirical data and apply ANN-based classification for the adoption of the most appropriate wireless channel model.

4.1. The Artificial Neural Network

An ANN [21] is a massively parallel distributed processor made up of simple processing units (neurons), which has the ability to learn functional dependencies from data. It resembles the brain in two respects:

(i) Knowledge is acquired by the network from its environment through a learning process.

(ii) Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

The procedure used to perform the learning process is called a learning algorithm, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective [22].

Each neuron is a simple processing unit which receives some weighted data, sums them with a bias and calculates an output to be passed on (Figure 4.1). The function that the neuron uses to calculate the output is called the activation function, where the output is

$$o = f\left(i_1 W_1 + i_2 W_2 + i_3 W_3 + \ldots + i_n W_n + b\right) = f\left(\sum_{j=1}^{n} i_j w_j + b\right) \tag{4.1}$$

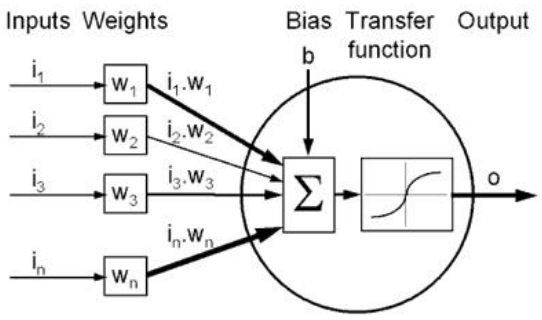


Figure 4.1. Graphical representation of an artificial neuron

Typically, activation functions are generally nonlinear having a squashing effect. Linear functions are limited because the output is simply proportional to the input. In Figure 4.2 some common activation functions are shown.

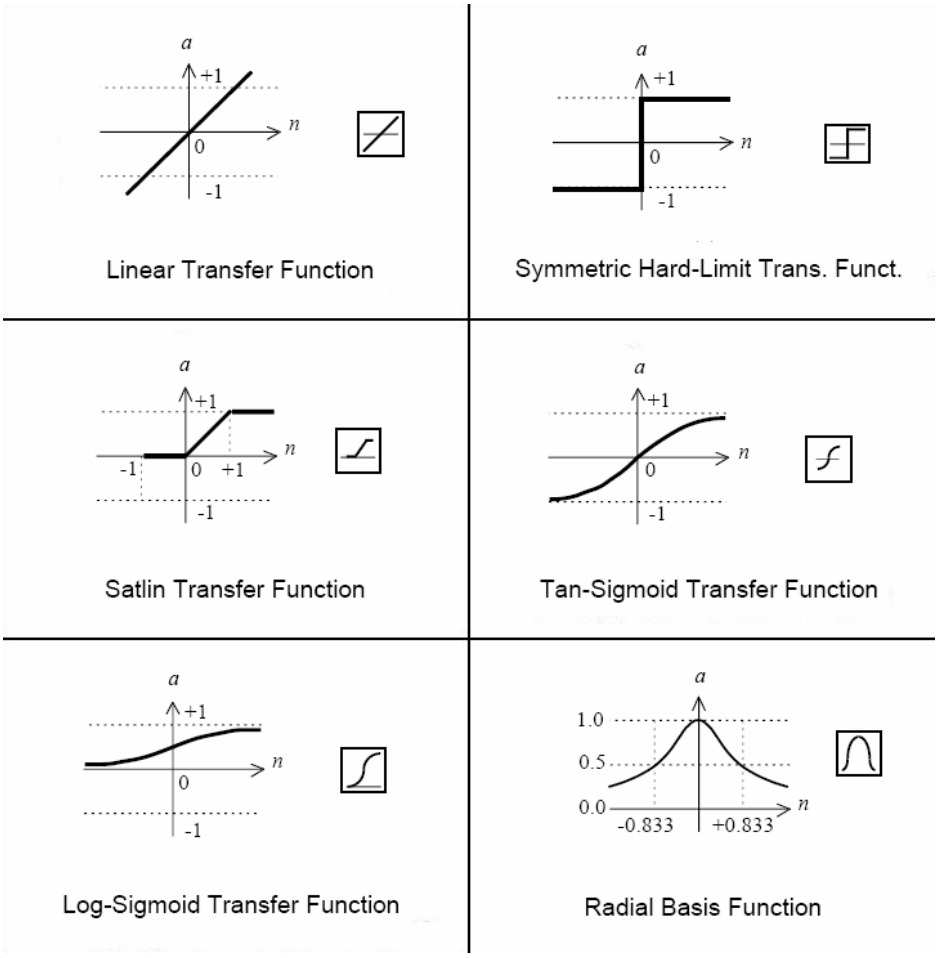


Figure 4.2. Common activation functions

4.1.1. Multilayer Perceptron

The manner in which the neurons of a neural network are structured is intimately linked with the learning algorithm used to train the network [22]. The most common architecture is the MLP. These networks are a feedforward network where the neurons are structured in one or more hidden layers. Each perceptron in one layer is connected to every perceptron on the next layer, hence information is constantly "fed forward" from one layer to the next.

A network can have several layers. Each layer has a weight matrix **W**, a bias vector **b**, and an output vector **a**. To distinguish between the weight matrices, output vectors, etc., for each of these layers, generally the number of the layer as a superscript to the variable of interest is appended. We can see the use of this layer notation in the three-layer network shown in Figure 4.3, and in the equations at the bottom of the figure.

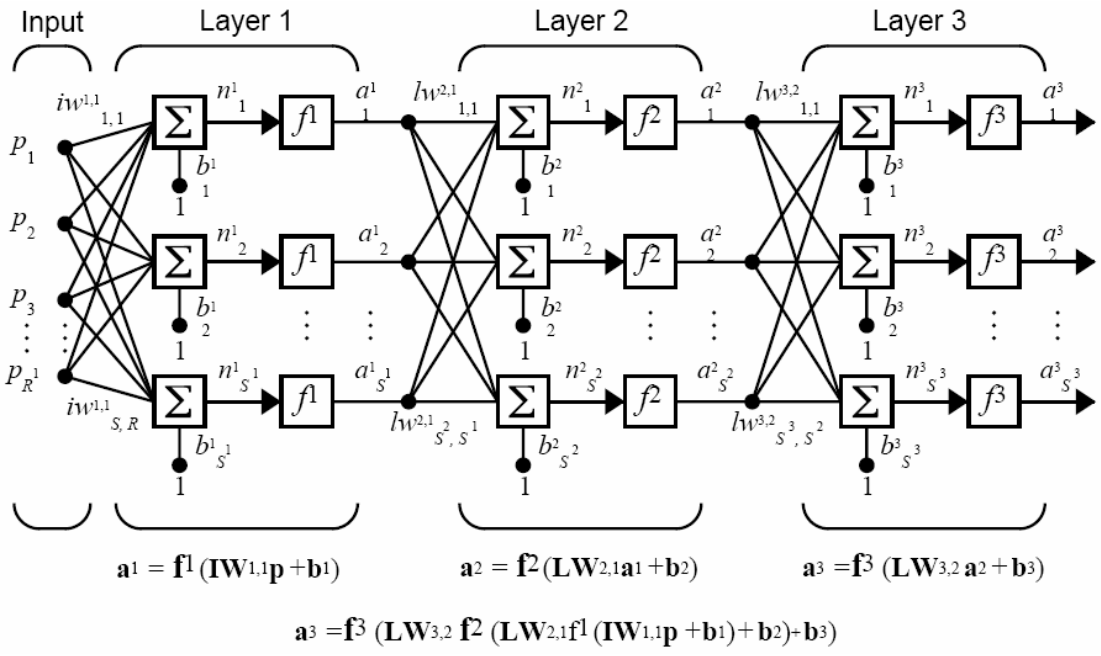


Figure 4.3. Multilayer perceptron network

The network shown above has R^1 inputs, S^1 neurons in the first layer, S^2 neurons in the second layer, etc. It is common for different layers to have different numbers of neurons. Here, a constant input 1 is fed to the biases for each neuron. The outputs of each

intermediate layer are the inputs to the following layer. Thus layer 2 can be analyzed as a one-layer network with S^1 inputs, S^2 neurons, and an $S^2 \times S^1$ weight matrix $\mathbf{W^2}$. The input to layer 2 is $\mathbf{a^1}$; the output is $\mathbf{a^2}$. Now that we have identified all the vectors and matrices of layer 2, we can treat it as a single-layer network on its own. This approach can be taken with any layer of the network.

The layers of a multilayer network play different roles. A layer that produces the network output is called an output layer. All other layers are called hidden layers. The three-layer network shown in Figure 4.3 has one output layer (layer 3) and two hidden layers (layer 1 and layer 2). By varying the number of nodes in the hidden layer, the number of layers, and the number of input and output nodes, one can classify points in arbitrary dimensional space into an arbitrary number of groups. However, the Hornik-Stinchcombe-White theorem [23] states that a layered ANN with two layers of neurons is sufficient to approximate as closely as desired any piecewise continuous map of a closed bounded subset of a finite-dimensional space into another finite-dimensional space, provided that there are sufficiently many neurons in the single hidden layer.

4.1.2. Back-propagation Algorithm

The network learns about the input through an interactive process of adjusting the weights and the bias. This process is called supervised learning and the algorithm used is the learning algorithm. One of the most common is the error back-propagation algorithm [21]. This algorithm is based on the error-correction learning rule, based on gradient descent in the error surface.

Basically, a set of cases, with the corresponding targets, is given to the network. The input data is entered into the network via the input layer and is processed through the layers. Then, the output is compared to the expected output (the targets) for that particular input. This results in an error value. This error value is backpropagated throughout the network, against the direction of the weights. The weights and bias are adjusted to make the actual response of the network move closer to the desired response in a statistical sense (Figure 4.4)

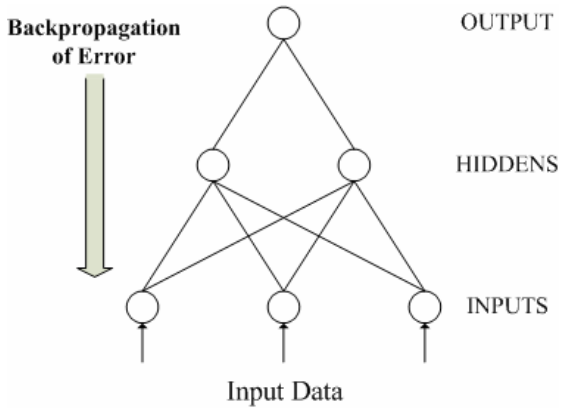


Figure 4.4. Back-propagation of error

According to Figure 4.5, the training algorithm of back-propagation can be described as follow:

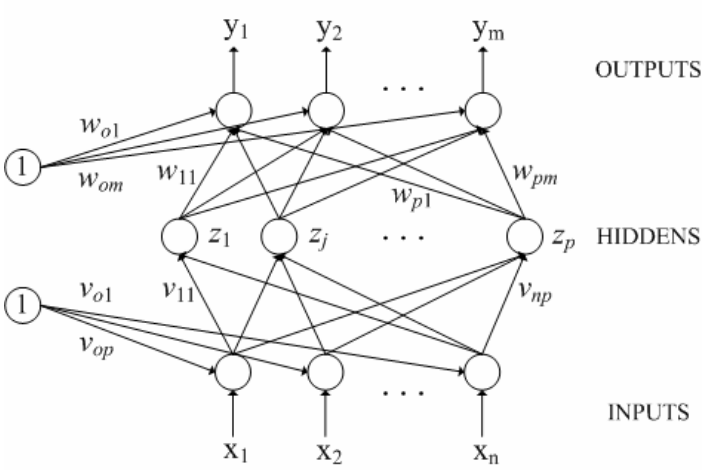


Figure 4.5. Back-propagation learning architecture in MLP

(i) Initialization of the weights:

Step 1: Initialize weight to small random values

Step 2: While stopping condition is false, do steps 3-10

Step 3: For each training pair do steps 4-9

(ii) Feedforward pass:

Step 4: Each input unit receives the input signal x_i and transmits this signals to all units in the hidden layer

Step 5: Each input unit receives the input signal z_j, $j = 1, ..., p$ sums its weighted input signals

$$z_{-inj} = v_{oj} + \sum_{i=1}^{n} x_i v_{ij} \tag{4.2}$$

applying activation function $Z_j = f\left(z_{-inj}\right)$ and sends this to all units in the output layer

Step 6: Each output unit y_k, $k = 1, ..., m$ sums its weighted input signals

$$y_{-ink} = w_{ok} + \sum_{j=1}^{p} z_j w_{jk} \tag{4.3}$$

and applies its activation function to calculate the output signal $Y_k = f\left(y_{-ink}\right)$

(iii) Backward pass:

Step 7: Each output unit y_k, $k = 1, ..., m$ receives a target pattern corresponding to an input pattern, error information term is calculated as

$$\delta_k = \left(t_k - y_k\right) f\left(y_{-ink}\right) \tag{4.4}$$

Step 8: Each hidden unit z_j, $j = 1, ..., p$ sums its delta inputs from units in the layer above

$$\delta_{-inj} = \sum_{k=1}^{m} \delta_j w_{jk} \tag{4.5}$$

The error information term is calculated as

$$\delta_j = \delta_{-inj} f\left(z_{-inj}\right) \tag{4.6}$$

(iv) Updating Weight and Biases:

Step 9: Each output unit y_k =1, ..., m updates its bias and weights (j = 0,..., p). The weight correction term is given by

$$\Delta W_{jk} = \alpha \delta_k z_j \tag{4.7}$$

and the bias correction term is given by

$$\Delta W_{ok} = \alpha \delta_k \tag{4.8}$$

Therefore, $W_{jk}\left(new\right) = W_{jk}\left(old\right) + \Delta W_{jk}, W_{ok}(new) = W_{ok}(old) + \Delta W_{ok}$. Each hidden unit z_j, j = 1, ..., p updates its bias and weights (i = 0,...,n).

The weight correction term

$$\Delta V_{ij} = \alpha \delta_j x_i \tag{4.9}$$

The weight correction term

$$\Delta V_{oj} = \alpha \delta_j \tag{4.10}$$

Therefore, $V_{ij}\left(new\right) = V_{ij}\left(old\right) + \Delta V_{ij}, V_{oj}(new) = V_{oj}(old) + \Delta V_{oj}$

Step 10: Test the stopping condition, the stopping condition may be the minimization of the errors, number of iterations etc.

4.2. The Decision Tree

A decision tree [24] is a tree whose internal nodes are tests on input patterns and whose leaf nodes are categories of patterns. An example is shown in Figure 4.6. A DT assigns a class number (or output) to an input pattern by filtering the pattern down through the tests in the tree. Each test has mutually exclusive and exhaustive outcomes.

For example, test T_2 in the tree of Fig 6.1 has three outcomes; the leftmost one assigns the input pattern to class 3, the middle one sends the input pattern down to test T_4 and the rightmost one assigns the pattern to class 1.

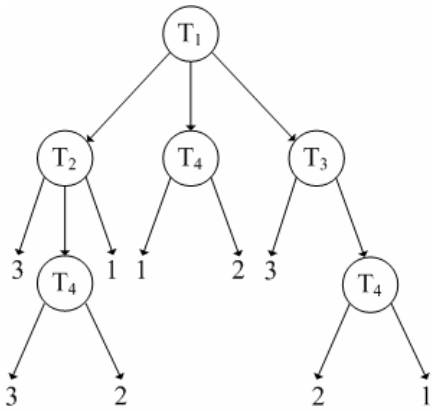


Figure 4.6. A decision tree

There are several dimensions along which DTs might differ:

(i) The tests might be *multivariate* (testing on several features of the input at once) or *univariate* (testing on only one of the features).

(ii) The tests might have two outcomes or more than two. (If all of the tests have two outcomes, we have a *binary DT*.)

(iii) The features or attributes might be categorical or numeric. (Binary-valued ones can be regarded as either.)

(iv) We might have two classes or more than two. If we have two classes and binary inputs, the tree implements a Boolean function, and is called a *Boolean DT*.

Classification trees have become increasingly important due to their conceptual simplicity and computational efficiency [25]. A DT classifier has a simple form which can be compactly stored and that efficiently classifies new data. DT classifiers can perform automatic feature selection and complexity reduction, and their tree structure provides easily understandable and interpretable information regarding the predictive or generalisation ability of the classification.

A tree is composed of a root node containing all the data, a set of internal nodes (splits), and a set of terminal nodes (leaves). Each node in a DT has only one parent node and two or more descendent node as shown in Figure 4.6. A dataset is classified by moving down the tree and sequentially subdividing it according to the decision framework defined by the tree until leaf is reached.

4.2.1. Decision Tree Building

To construct a classification tree by heuristic approach, it is assumed that a dataset consisting of feature vectors and their corresponding class labels are available. The features are identified based on problem specific knowledge. The DT is then constructed by recursively partitioning a dataset into purer, more homogenous subsets on the basis of a set of tests applied to one or more attribute values at each branch or node in the tree.

This procedure involves three steps: splitting nodes, determining which nodes are terminal nodes, and assigning class label to terminal nodes. The assignment of class labels to terminal nodes is straightforward: labels are assigned based on a majority vote or a weighted vote when it is assumed that certain classes are more likely than others.

DT is built through a binary recursive partitioning. This is a process of splitting the data into two partitions, left child or right child. This process is done recursively on each node until stopping criteria are met. At each node, one of the predictor variables is chosen as the splitting criteria. In addition, the range of the variable is specified to decide which

data goes to the left child and which data goes to the right child. In the case of continuous predictor, the split will be based on operators such as *greater than*. In the case of categorical predictor, the split will explicitly specify which category goes into which child node.

4.2.1.1. Splitting Node. At every node, each predictor variable is evaluated to see how well it could divide the data into two groups. The quality of the split can be calculated by the impurity of the child nodes. The lower the impurity of the resulting child nodes, the better the split.

The impurity of a node can be measured with the Entropy approach. In entropy approach, the impurity of the node is calculated by a weighted summation of the entropy of each child. The weight of a child node is the number of records it contains divided by total number of records in the parent node. The category which minimizes the misclassification cost for all rows in the node is chosen as the category to assign to the node. If all rows and categories are weighted the same, this implies majority rule in which each node will be assigned category which has the most number of rows in the node. For internal node, this does not matter much, as the predicted category will be predicted on leaf nodes.

4.2.1.2. Stopping Criterion. The DT built based on the method above might produce a very large tree. This very large tree could in fact works very well with the training dataset, 100 per cent accuracy is possible. But there are some disadvantages of having a large tree:

- It is computationally expensive
- It is difficult to interpret
- It might not generalize well to unseen data

The last item is unexpected, as the purpose of building DT is not to predict the training data set itself, but instead to predict a new data. A largely complex tree would allow each leaf node corresponds to one training data, which merely memorizing all the data, not generalizing. This would cause the tree to perform very badly in unseen data. Thus, it is beneficial to force a limit on how much the tree can grow.

There are two simple criteria to limit the growth of the tree:

(i) Minimum node size to split: This criterion will node split a node if the number of rows in the node is less than a certain threshold.

(ii) Maximum tree depth: This criterion will stop the node splitting that cause the depth of the tree to increase more than a certain limit.

4.2.1.3. Tree Pruning. The stopping criteria mentioned previously is a forward pruning. It is applied during the generation of the DT. However, it is often that the tree generated with forward pruning is not optimal in size.

Another more sophisticated approach is to use backward pruning. This means that initially, the tree can grow very large (with a more relaxed stopping criteria). In the next stage, this large tree is pruned to its optimal size. The optimal tree size is determined by a sophisticated algorithm based on validation of different tree size.

4.2.1.4. Cross Validation. The problem with the current method which builds the model from all training samples is that there is no indication on how well the resulting DT performs on an unseen data.

One approach to overcome this is to split the input data set into two categories: training data set and testing data set. The model building stage will only use the data in training data set. Once the model is built, it can then be tested using test data set. This method allows us to evaluate the accuracy of the newly built model.

There are three cross validation methods:

(i) Hold-out method: Divide the data set into training and testing set. Build the DT model using the training set and test the model using the testing set. This method is slightly better than residual (no testing) method. The problem with this method is that the testing result depends very much on which samples are chosen to be in the testing set.

(ii) K-fold cross validation: Divide the data set into k disjoint subsets. Repeat the hold-out method *k* times. Each time, *k-1* subsets are used as training set to build the DT and the other one subset is used to test the model. This reduces the variance problem of hold out method but increases the computation time by *k* times.

(iii) Leave one out validation: This is similar to K-fold cross validation with *K* is equal to *N* (the number of data).

4.2.1.5. Finding the Optimal Tree Size. As mentioned earlier, it is desirable to prune the tree to the optimal size. The optimal tree size can be determined by taking the average error of each tree size from each of the cross validation result.

Cross validation cost is the average error rate for a particular tree size. The optimal tree size is the tree size which produces minimum cross validation cost. Typically, as the tree size increases, the error decreases. This occurs up to a turning point where the error increases as the tree size increases. Once the optimal tree size is found, the overly large tree built in the initial stage will be pruned until it reaches the optimal tree size. The prune is done in stepwise fashion. At each step the least significant variable is removed.

5. ENVIRONMENT AWARE LOCATION ESTIMATION (EARBALE)

Since signal propagation characteristics are not the same, or even not similar for different kind of environments, it is improper to use a single generic signal attenuation model for all cases in the localization process. Even if we are dealing with a metropolitan city, it will very likely have areas that possess suburban or even rural signal propagation and path loss characteristics. Therefore, if we can estimate the environment type in which the MS resides, that will provide us valuable information to be used in the localization process.

EARBALE method uses preprocessing and dimensionality reduction via DT on empirical data collected in diverse geographic areas and applies ANN based pattern recognition for the adoption of the most appropriate wireless channel model (Figure 5.1). By this way, our trained ANN is capable of identifying the environment of an input measurement as urban, suburban, or rural. In this chapter, EARBALE method is explained in detail together with its architecture, phases, and key functions.

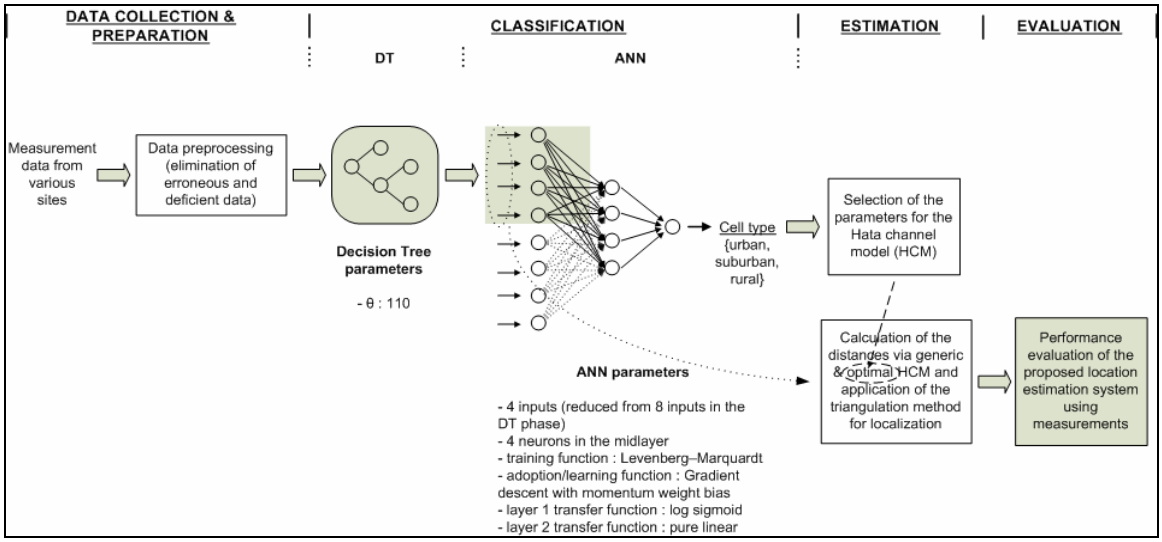


Figure 5.1. EARBALE architecture

5.1. Data Collection and Preparation Phase

For the location estimation system, raw measurement data have been gathered using a cell phone and a GPS device connected to a laptop running Ericsson TEMS™ Investigator software [26] as depicted in Figure 5.2. The data were collected in various locations of Istanbul with different environmental characteristics. Istanbul has a very complex and formidable geography for location estimation, having diverse terrain elements such as hills and valleys with dense settlement, the Bosporus Strait that separates two major parts of the city, the Princess Islands in south, and suburban and rural regions mainly located in north. These collected data were preprocessed to eliminate the erroneous and deficient data such as missing RSS levels or CGIs.

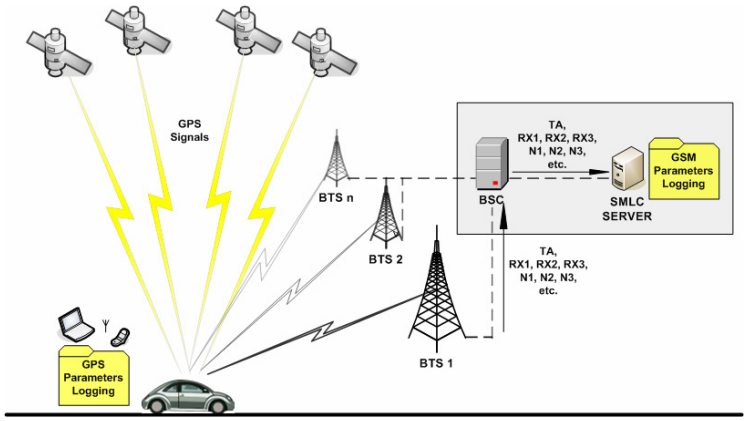


Figure 5.2. Field measurement process

The important issue with the collected experimental data is that the composition and measurement variables are governed by the signalling and network capabilities defined in GSM standards. Additionally, the GPS localization errors are not considered, which imply that the GPS outputs are considered as the exact mobile positions.

Data collection is done as a two-party collaborational process (Figure 5.3). While making drive-tests and recording the MS coordinates (via GPS) together with time stamps, the SMLC in the GSM network was recording the other required parameters such as TA, Rx values, serving BTS ID, neighboring BTS IDs, etc. sent by the MS together with, again,

time stamps. Afterwards, these two sources are fused to get a normalized and proper dataset. In order to eliminate some intrinsic problems (sticking with the serving BTS being the most important one) in continuous call-based localization scheme, we have switched to the trigger-based scheme. In trigger-based scheme, the MS used in the drive tests is not active all the time. Instead, it is activated periodically. In our case, the trigger interval is chosen to be one minute in the network side.

The eight input dimensions that will help us identify the environment at the beginning are as follows:

- $TA \equiv$ Timing advance parameter of the mobile station
- $Rx \equiv$ RSS parameter from the serving station
- $Rx_i \equiv$ RSS parameter from the i^{th} neighbor station, I $\in$ {1,...,3}
- $D_j \equiv$ Distance from the j^{th} neighbor station to the serving station, j $\in$ {1,...,3}

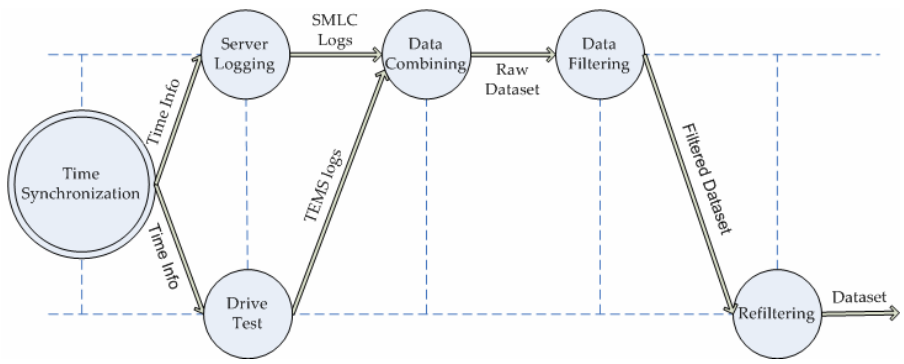


Figure 5.3. Data flow diagram for dataset acquisition

We use data from three out of five or six neighbors that can be provided by the GSM infrastructure. Weaker neighbors are eliminated as they are usually outliers (NLOS stations) so that we have first three strongest neighbors and their associated distance values. Also, azimuth values are not used (in contrast to our previous works [27, 28]) as environmental identifiers since they do not have significant contribution as separate inputs. Taking them into consideration as relative arguments with respect to each other in a way may work but for the time being that possibility has not been evaluated.

In summary, each proper measurement data is classified into the type of cell it belongs to in order to be used in the dimensionality reduction and the training phases of EARBALE. For this, all of the measurements in the dataset are classified as urban, suburban, or rural by examining them on Google Earth™ software [29]. To analyze the measurement results in detail, Google Earth KML (*Keyhole Markup Language*) files are created including CGIs, TA, serving BTS Rx level, azimuths, neighbor BTS Rx levels, GPS coordinates and timestamp for every measurement. This allows the visualization of actual environment, the base stations, the estimated position and actual position (GPS data) in a single setting and helps the interpretation of experimental results (Figure 5.4).

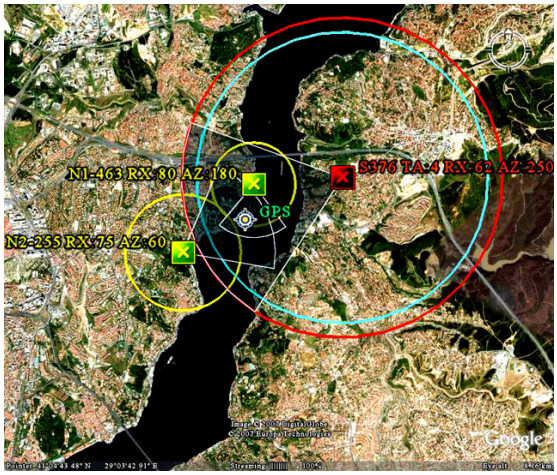


Figure 5.4. Dataset examination process in Google Earth™ software

5.2. ANN-based Classification Phase

Having a clean dataset with about 400 accurate measurements after the preprocessing phase, we move on to the machine learning based phase of EARBALE. The aim of this phase is to construct an ANN which is capable of identifying the environmental characteristics of the site where the MS resides so that the most suitable channel model for that site is used in the localization process. This section consists of the dimensionality reduction and the training subphases.

5.2.1. Dimensionality Reduction Subphase

We have eight inputs (eight-dimensions of data) for the training process that will be used for estimating the environment of a given measurement. These inputs are TA, Rx, Rx_1, Rx_2, Rx_3, d_1, d_2, and d_3. However, since these variables, especially distance values, will imply memory and computational load to the locationing system, a way to reduce the dimensionality to a lower degree is looked for. But, while doing that, we should not leave the classification accuracy behind. In other words, we look for the minimum input subset that can successfully classify the dataset measurements as we use all of the eight dimensions. For that purpose, we use the DT structure and find the minimum subset that give the best classification result by inspecting the relationship between the minimum split criterion value (Q) and the MSE (Figure 5.5).

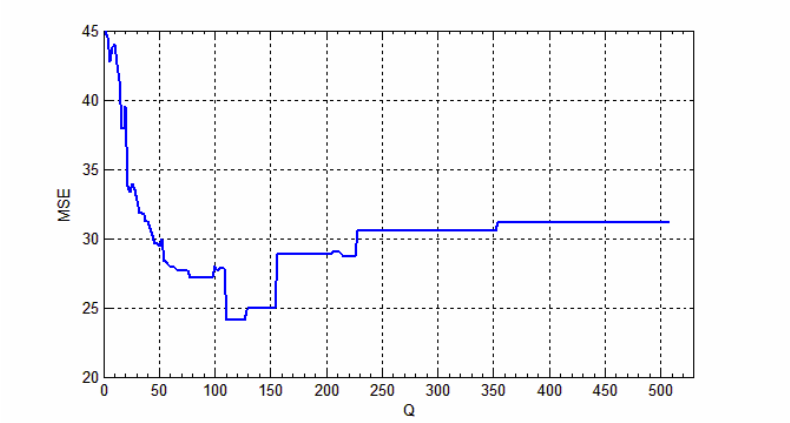


Figure 5.5. Effect of using the minimum split criterion on MSE

Q is an integer value that defines the minimum number of observations (measurements) that impure nodes must have to be split in the DT structure. If we choose a very small Q, it indicates an overfitting to the training data (memorizing) and the classification accuracy of the resulting DT will be low on the testing (validation) data. If we choose a very large Q, it indicates an underfitting to the data at hand and the classification error on the testing data will be again high. Between these extreme values, there is a Q' interval (or value) that gives the minimum error on the testing data. This interval was found to be [110, 130] for our dataset as can be seen in Figure 5.5. By selecting the Q value as 110, the DT is constructed as shown in Figure 5.6.

The resultant DT structure briefly tells us that by looking at TA and neighbor-serving station distances (d_1, d_2, and d_3), we can classify a given measurement as urban, suburban, or rural successfully as if we are using all of the eight input dimensions. This enables a simpler and more efficient classification module for EARBALE.

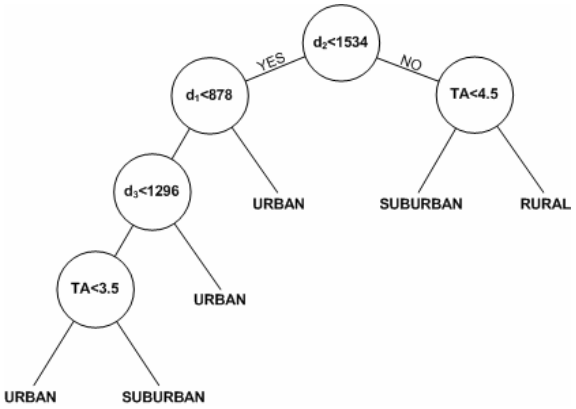


Figure 5.6. Resulting DT for feature selection (for *Q'=110*)

5.2.2. Training Subphase

After the dimensionality reduction phase via DT, we move on to the next step, ANN training. Neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. The structure of the three-layer ANN we use in EARBALE is shown in Figure 5.7.

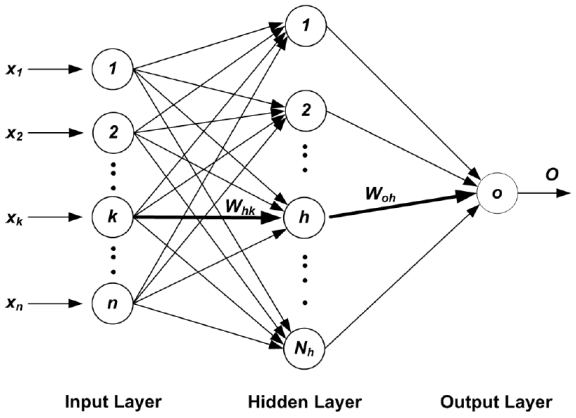


Figure 5.7. Three-layer ANN model used in EARBALE

The number of neurons in the input layer is equal to the number of input attributes given to the system, that is four in our case (*TA, d_1*, *d_2*, and *d_3*). The one neuron in the output layer gives the output *O,* that is the environment. The number of neurons in the hidden layer is four in our case as it gives the best results for our dataset.

Multiple-layer networks are quite powerful. A network of two layers, where the first layer is sigmoid and the second layer is linear, can be trained to approximate any function (with a finite number of discontinuities) arbitrarily well.

Our activation function for the neurons in the hidden and output layers is

$$f(y) = \frac{1}{1+e^{-y}} \tag{5.1}$$

where y is the input of the activation function. The input of the h^{th} neuron of the hidden layer is

$$NET_h = \sum_{k=1}^{n} W_{hk} \times x_k \tag{5.2}$$

where n is the number of input neurons, W_{hk} is the connection weight between the k^{th} neuron in the input layer and the h^{th} neuron in the hidden layer, and x_k is the input of the k^{th} neuron in the input layer. The output of the h^{th} neuron of the hidden layer is

$$OUT_h = f\left(NET_h\right) \tag{5.3}$$

The input of the output neuron is

$$NET_o = \sum_{h=1}^{N_h} W_{oh} \times OUT_h \tag{5.4}$$

where N_h is the number of neurons in the hidden layer and W_{oh} is the weight of connection between the h^{th} neuron in the hidden layer and the only one neuron in the output layer.

Then, the output of the ANN is calculated as

$$O = f(NET_o) \tag{5.5}$$

The details of the ANN learning process by back-propagation algorithm can be found in Chapter 4. Below, Figure 5.8 shows us the classification performance of our trained ANN whose accuracy is 75.8%. As we see, most of the misclassifications occur in the suburban class, which is the transition class. The reason of the worse classification accuracy for suburban areas is that there is no strict distinction between urban and suburban areas.

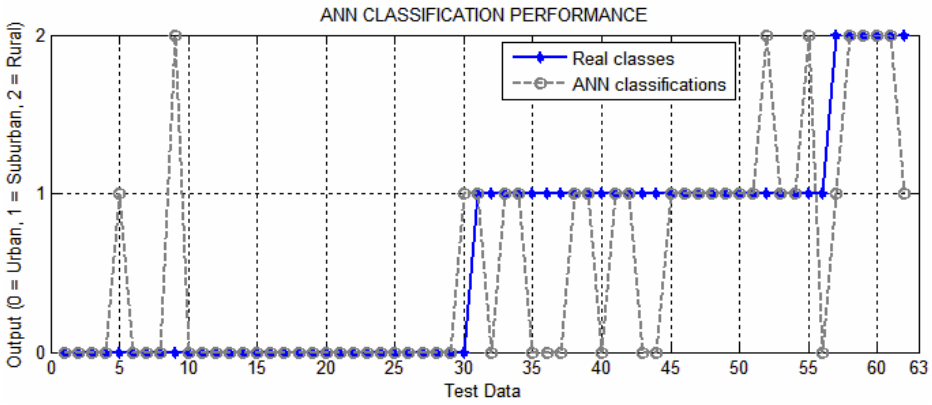


Figure 5.8. ANN Classification Performance

5.3. Estimation (Triangulation) Phase

In this phase, coordinates, antenna heights, antenna sector widths and azimuth angles of serving and neighbor BTSs are selected from the GSM operator's database in real-time. The positioning circles for each estimation are determined using these data together with the signal strength values and the channel model corresponding to the estimated environment in the previous ANN-based classifier phase.

In Figure 5.9, triangulation method for location estimation is elucidated with BTSs, positioning circles, cell sectors (azimuth angle + beam width) and the corresponding intersections. Knowing the coordinates, azimuths and beam witdhs of the BTSs together

with the radius values of the signal propagation circles acquired by the Hata's suitable path loss formulae, the intersection the candidate coordinates that the MS may be located can be found using the least squares method as described in Chapter 3.

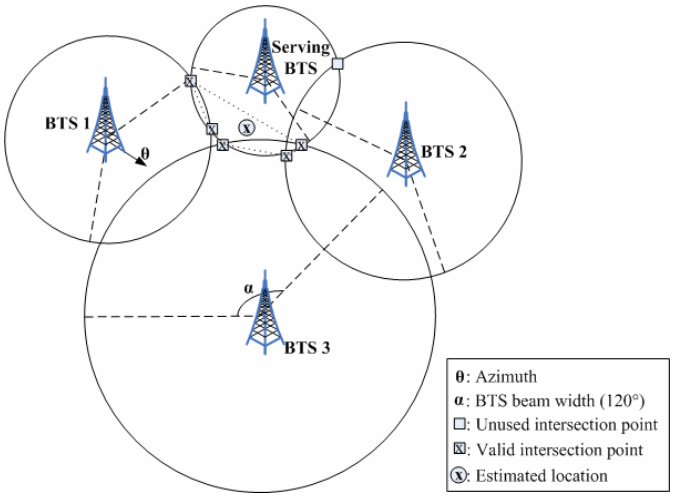


Figure 5.9. Location estimation via circle intersection using RSS values from serving and neighbor BTSs

After calculating all the circles, candidate intersection points that are not in the effective area of the serving BTS are eliminated. After this elimination, the coordinates of the MS is calculated as below:

$$lat_e = \frac{1}{n}\sum_{i=1}^{n} lat_i = \frac{1}{n}\left(lat_1 + ... + lat_n\right), \tag{5.6}$$

$$lon_e = \frac{1}{n}\sum_{i=1}^{n} lon_i = \frac{1}{n}\left(lon_1 + ... + lon_n\right), \tag{5.7}$$

where

- lat_e – latitude of the estimated point
- lon_e – longitude of the estimated point

- lat_i – latitude of the i^{th} of n intersection points
- lon_i – longitude of the i^{th} of n intersection points

In case of lack of network data due to any problem in GSM network or lack of intersection points, the system performs a fall-back and uses single BTS (i.e. serving BTS) based algorithm (Figure 5.10) when the minimum operating conditions cannot be satisfied. In this case, location of the MS is estimated as the midpoint between the Hata and TA circles along the azimuth angle.

TA value is an integer between 0 and 63 and each step represents an advance of one symbol period, which is approximately 3.69 microseconds. This means that the sensitivity of the TA attribute is about 550 meters. Hence, the radius of a TA circle, r_{TA} is found by the following simple formula:

$$r_{TA} = (TA + 1) \times 550 \tag{5.8}$$

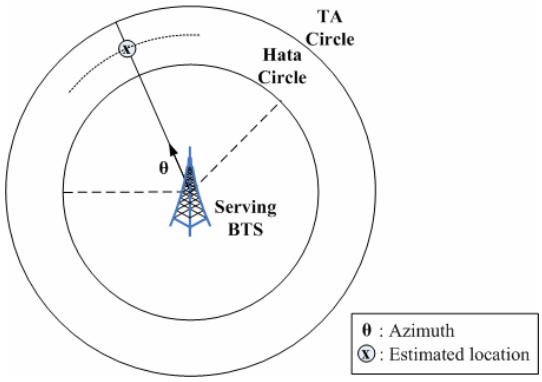


Figure 5.10. Singular BTS algorithm used as fall-back

5.4. Evaluation Phase

The dataset consists of the listed variables and the exact position of the user equipment determined using GPS. Therefore when any location estimation algorithm is run on the measurement, the actual position via GPS can be used to evaluate the positioning

performance just by calculating the Euclidean distance ε between the estimated and the actual positions. In other words,

$$\varepsilon = \sqrt{\left(lat_{GPS} - lat_e\right)^2 + \left(lon_{GPS} - lon_e\right)^2} \tag{5.9}$$

There are also other performance metrics such as maximum error and standard deviation of the localization error. The performance evaluation is described in detail in the following section.

6. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The general distribution of subscribers' TA in a location area is shown in Figure 6.1. Since this provides an idea for a realistic distribution of the cell environment property, our experimental data closely follow the same distribution, which is also elucidated in Figure 6.1. Our experimental data comprise of about 400 measurements in various environments. We have segmented the measurement data into 5 subsets, picked each one and used the remaining part for training (5-fold cross validation). Thus, we have performed five runs with different validation sets and averaged the results.

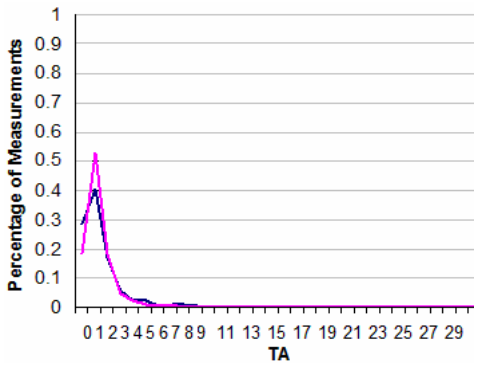
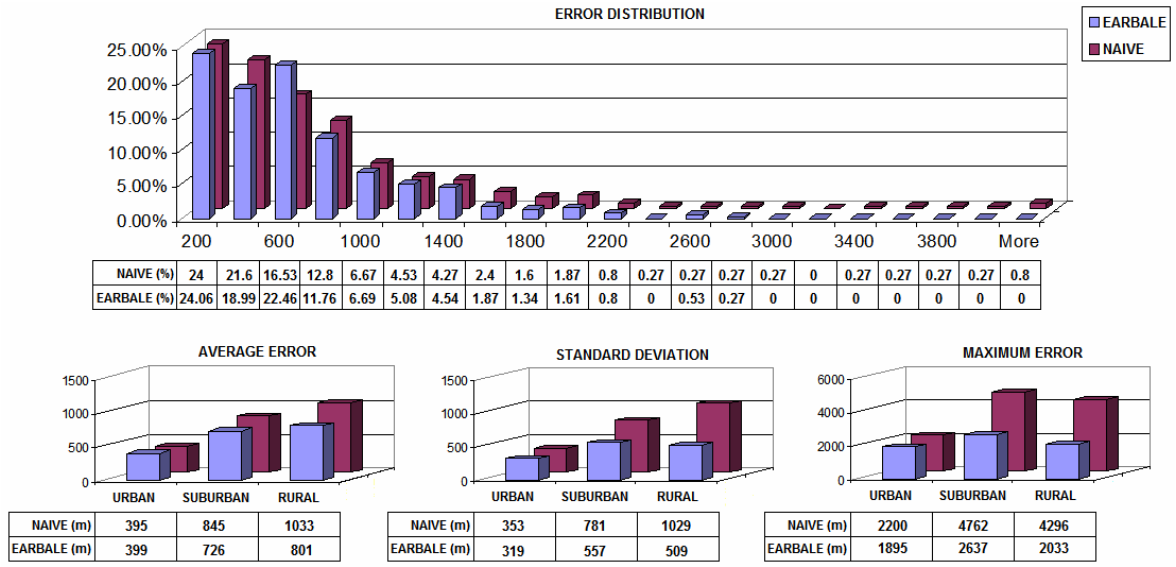


Figure 6.1. Subscribers' TA distribution in real GSM network (measured for 8016 subscribers in December, 2006) and in the dataset (dark line: dataset TA distribution, light line: GSM TA distribution)

In Figure 6.2, we present the detailed performance evaluation results in terms of error distribution, average error, maximum error, and standard deviation. In total, EARBALE reduced the average error from 642 meters to 573 meters (10.75% improvement), standard deviation from 689 meters to 481 meters (30.19% improvement), and maximum error from 4762 meters to 2638 meters (44.6% improvement). Apart from these numerical improvements, there is also one more important issue to emphasize that higher errors were mostly suppressed.



NAIVE (%)	24	21.6	16.53	12.8	6.67	4.53	4.27	2.4	1.6	1.87	0.8	0.27	0.27	0.27	0.27	0	0.27	0.27	0.27	0.27	0.8
EARBALE (%)	24.06	18.99	22.46	11.76	6.69	5.08	4.54	1.87	1.34	1.61	0.8	0	0.53	0.27	0	0	0	0	0	0	0

NAIVE (m)	395	845	1033
EARBALE (m)	399	726	801

NAIVE (m)	353	781	1029
EARBALE (m)	319	557	509

NAIVE (m)	2200	4762	4296
EARBALE (m)	1895	2637	2033

Figure 6.2. Performance evaluation results

When we observe the results in terms of environment types, we see that EARBALE also enhances the localization performance in all criteria. The average error for the urban environment is about the same for both naive and EARBALE methods because naive method already uses Hata's urban channel model for all cases since it has not the capability to distinguish environment types.

7. CONCLUSIONS

Location awareness is important to many pervasive computing applications. A fundamental problem in this context is location estimation, which is the estimation of a person's location from a stream of available data. This ability will allow the facilitation of novel location-aware services, which in turn provide better and richer service experience to the subscribers of wireless network operators. These network-based services, integrating a derived estimate of a mobile device's location or position with other information so as to provide added value to the user, are denoted as LBS.

However, accuracy is a crucial issue for current location estimation systems. Although the RSS method may not be as precise as some other localization methods in the literature, it is a widely used technique due to its ease of implementation on any cellular network since it does not require any changes to the existing handsets and network structures. In RSS based positioning techniques, we state that it is improper to use a single generic signal attenuation model for all cases in the localization process because that model may not suit to all of the cases. Even if we are dealing with a metropolitan city, it will very likely have areas that possess suburban or even rural path loss characteristics. For this reason, in this study we have tried to find a new methodology to identify the environmental characteristics of a given measurement so that the most suitable channel model for that environment is used in the localization process.

In the literature, the propagation characteristics of the environments have been modeled, but there is no such study that estimates the environmental characteristics of a measurement and utilizes this information for fine-tuning and enhancing RSS based distance calculations. The key feature of the developed method is its capability to estimate the environment of the mobile user as either urban, suburban or rural using machine learning techniques and to utilize this information for enhancing RSS based distance calculations.

We have presented our EARBALE scheme in all of its phases and the performance evaluation results in detail. We have shown that the proposed method enhances the generic algorithm in various criteria, including average error, maximum error, and standard deviation. The key feature is that EARBALE can be integrated onto any RSS-based location estimation scheme with minimal cost and impact on the legacy infrastructure. It is relatively lightweight and necessitates only a small amount of preliminary work and measurement activity.

REFERENCES

1. Rappaport, T. S., J. H. Reed and B. D. Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Communications Magazine*, vol. 34, issue 10, pp. 33-41, 1996.

2. Adusei, I. K., K. Kyamakyaand and F. Erbas, "Location-based Services: Advances and Challenges," in *Proc. Canadian Conf. on Electrical and Computer Engineering*, vol. 1, pp. 1-7, 2004.

3. Wann, C. D. and M. H. Lin, "Location Estimation with Data Fusion for Wireless Location Systems," in *Proc. the International Conference on Networking, Sensing & Control*, 2004.

4. Drane, C., M. Macnaughtan and C. Scott, "Positioning GSM telephones," *IEEE Communications Magazine*, vol. 36, issue 4, pp. 46-54, 59, 1998.

5. Jami, I., M. Ali and R. F. Ormondroyd, "Comparison of Methods of Locating and Tracking Cellular Mobiles," in *Proc. IEEE Colloquium on Novel Methods of Location and Tracking of Cellular Mobiles and Their System Applications*, pp. 1/1 - 1/6, 1999.

6. Zhao, Y., "Mobile Phone Location Determination and Its Impact on Intelligent Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, issue 1, pp. 55-64, 2000.

7. Spirito, M. A., S. Poykko and O. Knuuttila, "Experimental Performance of Methods to Estimate the Location of Legacy Handsets in GSM," in *Proc. IEEE Vehicular Technology Conference*, vol. 4, pp. 2716-2720, 2001.

8. Wann, C. and Y. Chen, "Position Tracking and Velocity Estimation for Mobile Positioning Systems," in *Proc. The 5th International Symposium on Wireless Personal Multimedia Communications*, vol. 1, pp. 310–314, 2002.

9. Cong, L. and W. Zhuang, "Hybrid TDOA/AOA Mobile User Location for Wideband CDMA Cellular Systems," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 439 – 447, 2002.

10. Venkatraman, S. and J. J. Caffery, "Hybrid TOA/AOA techniques for mobile location in non-line-of-sight environments," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Atlanta, USA, pp. 274-278, 2004.

11. Ng, K. J., S. K. C. Chan and K. K. H. Kan, "Providing Location Estimation within a Metropolitan Area based on a Mobile Phone Network," in *Proc. The 13th International Workshop on Database and Expert Systems Applications*, 2002.

12. Hata, M., "Empirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317-325, 1980.

13. Tonteri, T., *A Statistical Modeling Approach to Location Estimation*, M.S. Thesis, University of Helsinki, 2001.

14. Parsons, J. D., *The Mobile Radio Propagation Channel*, pp. 33–37, 2nd Edition, John Wiley & Sons, 2000.

15. Walfisch, J. and H. L. Bertoni, "A Theoretical Model of UHF Propagation in Urban Environments," *IEEE Trans. Antennas Propagat.*, vol.36, pp.1788-1796, 1988.

16. Thomas, P. A., S. M. Nabritt and M. A. Belkerdid, "Propagation Models Used in Wireless Communications System Design", in *Proc. IEEE Southeastcon*, 1998.

17. Laitinen, H., S. Ahonen, S. Kyriazakos, J. Lähteenmäki, R. Menolascino and S. Parkkila, *Cellular Network Optimisation Based on Mobile Location*, Technical Report, 2001.

18. Apaydın, G., *Comparison of Location-Estimation Techniques of GSM Phones with the Simulations*, M.S. Thesis, Boğaziçi University, 2003.

19. Sayed, A. H., A. Tarighat and N. Khajehnouri, "Network Based Wireless Location: Challenges Faced in Developing Techniques for Accurate Wireless Location Information," *IEEE Signal Processing Magazine*, Vol. 22, July 2005.

20. LaMance, J., J. Jarvinen and J. DeSalas, *Assisted GPS: A Low-Infrastructure Approach*, GPS World, 2002.

21. Oliveira, A., Neural Network Software Tool Development: Exploring Programming Language Options, Technical Report, 2006.

22. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 2nd Edition, 1998.

23. Looney, C., *Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists*, Oxford University Press, 1997.

24. Nilsson, N. J., *Introduction to Machine Learning*, pp. 81-82, 1996.

25. Pal, M. and P. M. Mather, "Decision Tree Based Classification of Remotely Sensed Data," in *Proc. Asian Conference on Remote Sensing*, 2001.

26. Ericsson TEMS™ Investigator, http://www.ericsson.com/solutions/tems/

27. Bayrak, O., C. Temizyurek, M. Barut, O. Turkyilmaz and G. Gur, "A Novel Mobile Positioning Algorithm Based on Environment Estimation," in *Proc. 4th Workshop on Positioning, Navigation and Communication*, 2007.

28. Bayrak, O., C. Temizyurek, M. Barut, O. Turkyilmaz and G. Gur, "Ortam Tahminine Dayalı Yeni Bir Mobil Konumlama Algoritması," *IEEE 15. Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, 2007.

29. GOOGLE EARTH™ Software, http://earth.google.com/