

LOCATION ESTIMATION BY FINGERPRINTING
IN CELLULAR NETWORKS

by

Ömer Faruk Kurt

B.S. in Computer Engineering, Koç 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
2009

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Assoc. Prof. Tuna Tuğcu for guiding through my thesis. Secondly, I want to thank for his support and patience throughout all my graduate study and on this thesis.

I also would like to thank Assoc. Prof. Fatih Alagöz for his support and patience during my graduate education.

Also, I would like to thank to TÜBİTAK for their financial support during my graduate education.

Finally, I would like to thank my father, Ekrem Kurt, my mother, Melike Kurt, my brother Engin Can Kurt, my relatives Şevket Akkaya, Habib Kurt, and my girlfriend Esra Güney for their endless support, tolerance, and love during my educational life.

ABSTRACT

LOCATION ESTIMATION BY FINGERPRINTING IN CELLULAR NETWORKS

Location based services have attracted too much attention recently due to the increasing number of new value-added services. Also, the requirement set by Federal Communications Commission (FCC) to locate emergency calls according to certain accuracy criteria has been the key factor for the development of location methods. Hence, location positioning has been a key research problem and numerous localization solutions have been proposed. These include technologies such as Cell ID, angle and time of arrival methods, statistical methods and fingerprinting methods. This paper proposes a fingerprinting based positioning method.

Location fingerprinting is a novel method that benefits from the previous signal measurements to find the location of a user. The previous signal measurements are stored into a database, called fingerprint database, to be used for fingerprint matching. Fingerprinting consists of four phases such as data collection, fingerprint filtering, fingerprint matching, and location estimation algorithms.

In this thesis, a new fingerprint matching algorithm is proposed to increase the accuracy of the location based services. This new fingerprint matching algorithm is compared with the most frequently used and well-known algorithms. The proposed algorithm does not require any hardware modifications. It does not require any modifications in the software on the mobile devices, either. By exploiting the measurements that are already being performed by the cellular network, the system can estimate the location of the mobile station using the RSS values of the MS. It is shown that the proposed algorithm gives better results than the compared algorithms in terms of accuracy.

ÖZET

PARMAK İZİ YÖNTEMİ İLE HÜCRESEL AĞLARDA KONUM BELİRLEME

Katma değerli yeni servislerin sayısının artması ile beraber konuma dayalı servislerin önemi artmaktadır. Acil aramaların nereden yapıldığının belirli bir hata payı içinde tespit edilmesini şart koşan Uluslararası İletişim Komisyonu da konuma dayalı servisler üzerinde anahtar bir etken olmuştur. Bu nedenle konum belirleme anahtar bir araştırma sorunu olmuş ve birçok sayıda konum belirleme çözümü sunulmuştur. Bu çözümlerden bazıları hücre numarası, varış zamanı ve açısı, istatistiksel ve parmak izi yöntemleridir. Bu tez parmak izi yöntemine dayalı bir konum belirleme yöntemi öne sürmektedir.

Parmak izine bağlı konum belirleme bir kullanıcının yerini bulabilmek için daha önce toplanmış sinyal ölçümlerinden yararlanır. Önceki sinyal ölçümleri sonradan sinyal izi eşleştirmede kullanılmak için bir veritabanına kaydedilir ve bu veritabanına sinyal izi veritabanı denir. Sinyal izine bağlı konumlandırma dört aşamadan oluşmaktadır, bunlar ölçüm toplama, sinyal izi filtreleme, sinyal izi eşleştirme ve konum hesaplamadır.

Bu tezde konuma dayalı servislerin hassaslığını artırmak için yeni bir sinyal izi eşleştirme yöntemi önerilmektedir. Bu yeni yöntem en çok bilinen ve en çok kullanılan yöntemler ile kıyaslanmıştır ve performansı bu algoritmalar ile karşılaştırılmıştır. Önerdiğimiz yöntem hücresel ağda veya cep telefonlarında fazladan bir donanım yada yazılıma ihtiyaç duymadan çalışabilmektedir. Sistem mobil kullanıcının sinyal değerlerini ve hücresel ağ tarafından yapılan ölçümleri kullanarak kullanıcının yerini tespit edebilmektedir. Bu tezde önerilen yöntemin karşılaştırılan yöntemlerden konum hassasiyeti açısından daha iyi sonuçlar verdiği gösterilmiştir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZET	v
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF SYMBOLS / ABBREVIATIONS	xi
1. INTRODUCTION	1
1.1. Organization of Thesis	4
2. RADIO WAVE PROPAGATION	5
2.1. Basic Principles	6
2.1.1. Path Loss	8
2.1.2. Attenuation	10
2.1.2.1. Free Space Loss	10
2.1.3. Absorption	11
2.1.4. Diffraction	11
2.1.5. Reflection	12
2.1.6. Refraction	13
2.2. Propagation Models	14
2.2.1. The Okumura-Hata Model	15
2.2.2. The Cost 231-Hata Model	18
2.2.3. The Walfisch-Bertoni Model	18
3. LOCATION ESTIMATION METHODS	20
3.1. Cell Global Identification and Timing Advance	21
3.2. Time of Arrival	22
3.3. Time Difference of Arrival	25
3.4. Angle of Arrival	27
3.5. Assisted GPS	29
3.6. Signal Strength	31
3.7. Fingerprinting	34

4. FINGERPRINTING METHODS.....	36
4.1. Fingerprint Matching Methods.....	36
4.1.1. GSM for Location Method.....	36
4.1.2. Euclidian and Manhattan Distance Method	37
4.2.1.1. Euclidian Distance	38
4.2.1.2. Manhattan Distance	38
4.1.3. DCM for Multi-System Location	38
4.1.4. Novel Cost Function for DCM.....	39
4.1.5. DCM by Using Average RSS and Standard Deviation	39
4.2. Location Estimation Methods	40
4.2.1. Nearest Neighbor (NN) method	40
4.2.2. Weighted k Nearest Neighbor (WkNN) method	41
5. LOCATION ESTIMATION BY FINGERPRINTING (LEF).....	42
5.1. Data Collection.....	42
5.2. Fingerprint Filtering.....	44
5.3. Fingerprint Matching Algorithm	44
5.3.1. LEF Fingerprint Matching Algorithm	45
5.3.2. LEF Training	45
5.4. Location Estimation Algorithms	46
6. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION	47
6.1. Dataset Analysis	47
6.1.1. Istanbul Dataset	47
6.1.2. Izmir Dataset	49
6.2. Threshold and Coefficient Calculation	52
6.3. Performance Comparison.....	52
6.4. Number of Nearest Neighbor Calculation	55
7. CONCLUSIONS	59
REFERENCES.....	60

LIST OF FIGURES

Figure 1.1. The elements of a cellular network	1
Figure 2.1. The radio wave propagation landscape.....	5
Figure 2.2. Power transmission elements	8
Figure 2.3. Diffraction.....	12
Figure 2.4. Two-ray ground reflection model.....	13
Figure 2.5. Refraction.....	14
Figure 2.6. The Walfisch-Bertoni propagation model	19
Figure 3.1. CGI and Timing Advance Method	22
Figure 3.2. TOA method.....	23
Figure 3.3. TDOA method.....	25
Figure 3.4. AOA method	28
Figure 3.5. A-GPS architecture.....	30
Figure 3.6. MS-BTS geometry.....	31
Figure 5.1. Field Measurement Process	43
Figure 6.1. Istanbul measurements.....	48

Figure 6.2 TA histogram of Istanbul measurements	48
Figure 6.3 Rx histogram of Istanbul measurements.....	49
Figure 6.4. Izmir measurements.....	50
Figure 6.5. TA histogram of Izmir measurements	50
Figure 6.6. Rx histogram of Izmir measurements.....	51
Figure 6.7. Average error in distance for Istanbul	53
Figure 6.8 CDF comparison for Istanbul Measurements	54
Figure 6.9. Average error in distance for Izmir	54
Figure 6.10 CDF comparison for Izmir	55

LIST OF TABLES

Table 6.1. Number of measurements for same serving BTS in Istanbul.....	49
Table 6.2. Number of measurements for same serving BTS in Izmir.....	51
Table 6.3. Threshold and Coefficient Values for two Measurement Sets.....	52
Table 6.4 Mean comparison for Istanbul.....	54
Table 6.5 Mean comparison for Izmir	55
Table 6.6. Optimum k value for Istanbul and Izmir.....	57

LIST OF SYMBOLS / ABBREVIATIONS

A-GPS	Assisted Global Positioning System
AOA	Angle of Arrival
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CGI	Cell Global Identification
DCM	Database Correlation Method
ED	Euclidean Distance
GPS	Global Positioning System
GSM	Global System for Mobile Communications
FSPL	Free Space Path Loss
LBS	Location Based Services
LEF	Location Estimation by Fingerprinting
LOS	Line of Sight
HLR	Home Location Register
MD	Manhattan Distance
MS	Mobile Station
MSC	Mobile Switching Center
MSL	Multi-System Location
NLOS	Non Line of Sight
PSTN	Public Switched Telephone Network
RSSI	Received Signal Strength Identifier
RT	Regression Tree
TA	Timing Advance
TDOA	Time Difference of Arrival
TOA	Time of Arrival
VLR	Visitor Location Register

1. INTRODUCTION

Location Based Services (LBS) are information services that make use of the location of mobile device. LBS provide a broad range of new applications to the cellular operators. Also with the information of the location, the importance and penetration of the current value added services are increased. Location based advertising, dating, games, media, billing, and emergency services are some of the new and current value added services that LBS have an impact on. Consequently, mobile positioning in wireless networks has been playing a crucial role in both research and industry [1, 2].

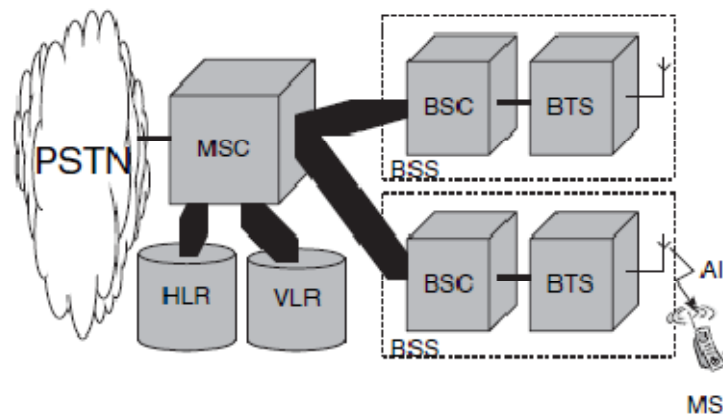


Figure 1.1. The elements of a cellular network

The location of the mobile can be found by using handset-based or network-based location estimation techniques. Handset based location estimation requires special equipment in the handsets. To be able to calculate its position independent of the network, handsets generally require GPS receivers [3]. GPS based positioning gives more accurate results than network based techniques because of the working mechanism of the GPS [40]. On the other hand, it has also some drawbacks. The cost of handsets with GPS receiver is high, and GPS receivers consume battery of the mobile equipments. As a result the penetration of handsets with GPS receivers is very low. Consequently, the adoption of LBS will take time due to the slow spread of new handsets. Last drawback is that GPS-based methods cannot be applied for indoor positioning. Handset-based methods that utilize RSSI

measurements rather than GPS are not practical since the handsets are not powerful enough for complex location calculations.

On the contrary, using network-based method for mobile positioning has many advantages over using the handset-based methods. The location of a handset can be calculated by this method even if it does not have a GPS receiver. So the adoption of LBS is expected to be faster than the handset-based methods. Furthermore, network based methods can be applied for both outdoor and indoor positioning. The only drawback of network-based method is its accuracy. So, research generally focuses on improving the accuracy of the network-based method.

In both handset-based and network-based positioning, a wide variety of techniques have been proposed to calculate the location of the mobile. These techniques use the measurements performed within the cellular environment. Some of the methods used in both kinds of positioning are satellite-based GPS technology, stand-alone GPS, Assisted GPS (A-GPS), Time of Arrival (TOA), Angle of Arrival (AOA), Time Difference of Arrival (TDOA), Received Signal Strength (RSS), and Database Correlation Method (DCM) [4, 5, 6, 7, 8].

To be able to calculate the position by GPS, a GPS receiver should be placed in the handsets. With the A-GPS method, placing a partial GPS receiver into the handsets is not sufficient. It also requires assistance servers to function properly [31]. Time of arrival (TOA) can be used with legacy handsets. However, specialized network elements Location Measurement Units (LMU) are needed in base stations. Angle of arrival (AOA) method requires antenna arrays at BTSs or terminals [9, 10]. Since putting this extra equipment at terminals is more expensive and complex than putting them at BTSs, antenna arrays are generally deployed at BTSs. In these methods, extra equipment is needed in handsets or cellular networks.

The received signal strength indicator (RSSI) based method makes use of the information measurable in cellular networks. Timing advance (TA), Rx level and Cell Global Identity (CGI) values are the information used by an RSSI-based method. TA value corresponds to the length of time which required for the transmission of a signal from the

base station (BS) to the mobile station (MS). In TDMA technology, sequential time slots are assigned to the users who share the same frequency. TA value also used to adjust the timeslot during which a user is allowed to communicate [32]. Rx levels are the strength of signals received by the MS. In GSM, MS measures one Rx value from serving BTS and six Rx value from the strongest neighbor BTSs. The value of Rx depends on the loci of the MS and the serving BTS. RF signals attenuate during transmission. The attenuation values are predicted by Okumura-Hata, Cost 231-Okumura, Walfish-Bertoni, etc. These methods are used to calculate the location of MS [11, 17, 18].

Received signal strength is the only common information in all kinds of handsets. Although RSS based methods have drawbacks due to signal attenuation, they are more applicable than the other methods. To maximize the accuracy of RSS-based methods, new approaches are proposed for location estimation [12].

Database correlation method can be applied to any cellular network [26]. Signal information taken from the MS is stored in the database. Stored signal information samples are called fingerprints. Fingerprints could include signal strengths, time delay, and other useful information. Fingerprints depend on the geographic location and the cellular network. The location estimation system is trained with the samples taken from the network. When the location of the MS is required, the MS sends its measurements to the network. The location estimation system decides the location of the MS by comparing the signal measurements of MS against the database.

The literature contains a wide range of cellular positioning technologies. Yet, none of them have proven superior in terms of accuracy for all environments worldwide [15, 26, 28, 36, 39, 40, 46, 47, 48, 49]. On the other hand, a positioning technology, first proposed in [26] known as database correlation method (DCM), has shown the potential for very high accuracy in different environments [26, 27, 28, 39, 49, 50].

In this thesis, we propose a location estimation method based on fingerprinting, namely Location Estimation by Fingerprinting (LEF) and evaluate its performance. We utilize signal measurements in a 900 MHz GSM network and real data measurements compiled from Istanbul and Izmir. LEF proposes a new fingerprint matching algorithm to

increase the accuracy of the fingerprint based location estimation techniques. We compare the new fingerprint matching algorithm with the most-frequently used and well-known matching algorithms and evaluate its performance. We make experiments with different sets of measurements to analyze the algorithm. We also divide the whole set of samples into subsets to observe the effect of the number of measurements under the same BTS.

1.1. Organization of Thesis

This thesis consists of five sections. Chapter 2 presents the basic principles of radio wave propagation and the artificial intelligence. In Chapter 3, location estimation methods are investigated. In Chapter 4, the methods of Fingerprinting are explained. In Chapter 5, our approach to location estimation is presented. Chapter 6 presents the evaluation and the results of this study. Finally in Chapter 7, the study is finalized with concluding remarks.

2. RADIO WAVE PROPAGATION

The propagation properties of a signal determine the accuracy of location estimation in network based methods. The propagation of the signal depends on the propagation environment such as urban or mountainous areas. Figure 2.1 shows a few of the many interactions between radio waves. Phenomena such as attenuation, diffraction, reflection, refraction, and scattering have important roles in location estimation since they affect the radio waves and the RSS values measured by the MS [13, 14, 15]. This chapter explains the theoretical aspects of radio wave propagation and some propagation models based on them.

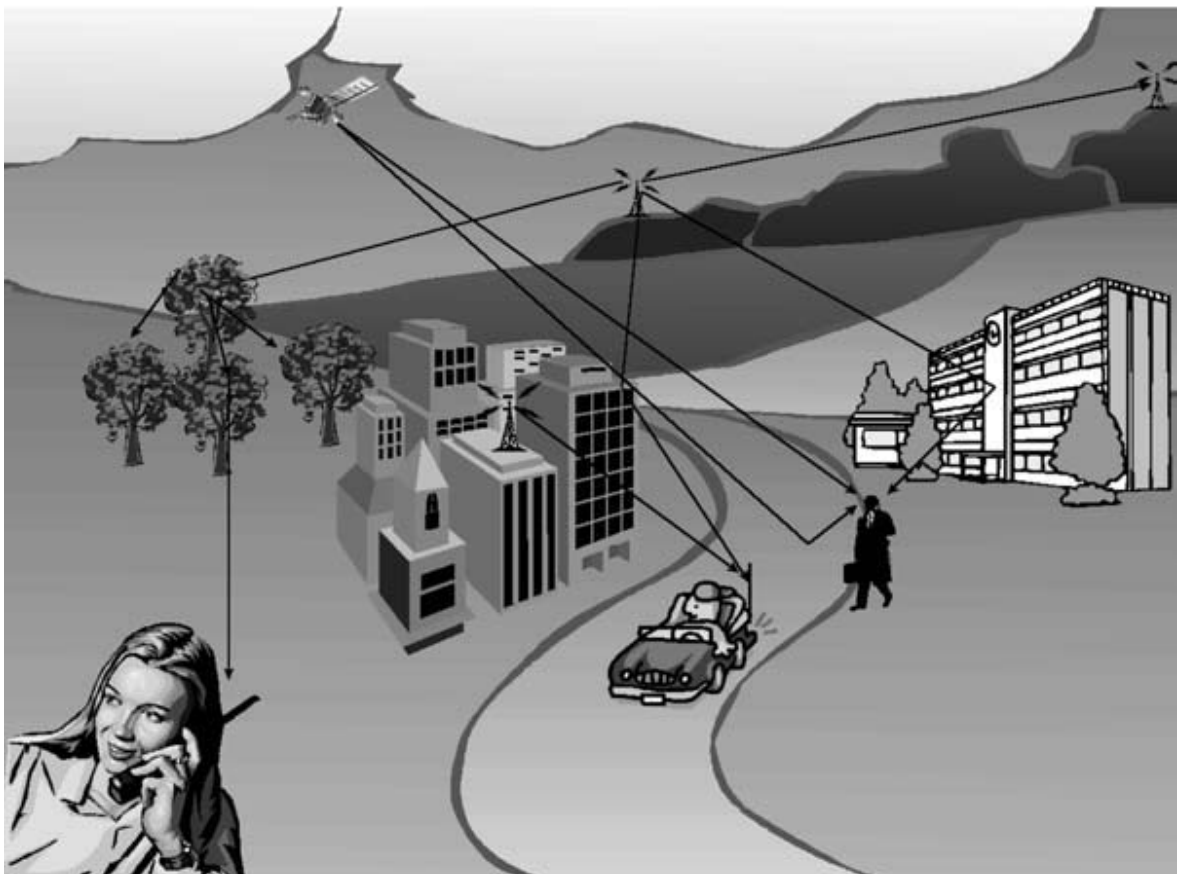


Figure 2.1. The radio wave propagation landscape

2.1. Basic Principles

Electromagnetic radiation has electric and magnetic field components. There are several types of electromagnetic radiation such as radio waves, microwaves, infrared radiation, etc. The frequency of the wave determines the type of electromagnetic radiation. Radio waves are the electromagnetic waves transmitted on the radio frequencies. Antennas are needed to radiate and receive electromagnetic energy. An antenna is a transducer designed to transmit or electromagnetic waves [13].

The radiation pattern of an antenna describes the relative strength of the radiated field in various directions from the antenna. Radiation intensity U is the power radiated from an antenna per unit solid angle. The plot of the radiation intensity defines the radiation pattern of the antenna [13]. U is calculated simply by multiplying the power density at a given distance by the square of the distance r , where S (W/m^2) is the power density

$$U = r^2 S. \quad (2.1)$$

An isotropic antenna is an idealized antenna which radiates equally in all directions. The power is spread over a sphere of radius r from the antenna. If P is the total power radiated by the antenna, the power density S (W/m^2) at r distance and in any direction is

$$S = \frac{P}{\text{area}} = \frac{P}{4\pi r^2}. \quad (2.2)$$

If r is within the far field, the effect of distance is removed from the radiation intensity. This ensures that radiation pattern is the same at all distances from the antenna. By combining equations 2.1 and 2.2, the radiation intensity is calculated as

$$U = r^2 S = \frac{P}{4\pi} \quad (2.3)$$

which is independent of r .

An antenna produces a power P_r watts in its terminating impedance of and is used to receive a wave with a power density S (W/m^2). P_r is the constant proportion of S defined by the effective aperture A_e . The effective aperture of the antenna in square meters is calculated as

$$A_e = \frac{S}{P_r}. \quad (2.4)$$

The aperture is described as the physical area of the antenna [41]. The effective aperture may often be much larger than the physical area, especially in the case of wire antennas. The effective aperture reduces as the efficiency of an antenna decreases [13]. The effective aperture is related to the antenna gain G as

$$G = \frac{4\pi}{\lambda^2} A_e. \quad (2.5)$$

If the power input to antenna A is P_t , then, accordance with (2.2), the power density incident on antenna B in is

$$S = \frac{P_t G_a}{4\pi r^2} \quad (2.6)$$

where G_a is the maximum gain of antenna A. Applying equation (2.2), the received power at the terminals of antenna B is

$$P_r = \frac{P_t G_a A_{eb}}{4\pi r^2} \quad (2.7)$$

where A_{eb} is the effective aperture of antenna B. When equation 2.5 is substituted for A_{eb} , the relation between the received and transmitted powers becomes

$$\frac{P_r}{P_t} = G_a G_b \left(\frac{\lambda}{4\pi r} \right)^2. \quad (2.8)$$

2.1.1. Path Loss

The transmitted power decreases on the way to the antenna due to several interactions such as attenuation, absorption, diffraction, reflection, and refraction. The path loss between a pair of antennas is the ratio of the transmitted power to the received power. It includes all possible elements of loss associated with interactions between the propagating wave and any objects between the transmit and the receiving antennas. The power appearing at the receiver input terminal is denoted as P_R and can then be expressed as

$$P_R = \frac{P_T G_T G_R}{L_T L L_R} \quad (2.9)$$

where the parameters are defined in Figure 2.2, with all gains G and losses L expressed as power ratios and powers expressed in watts.

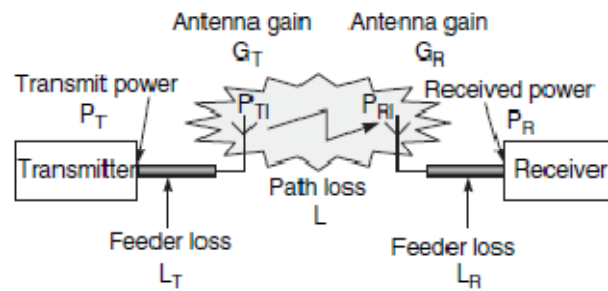


Figure 2.2. Power transmission elements

The antenna gain is defined as the ratio of the radiation intensity of an antenna in a given direction to the intensity that would be produced by an isotropic antenna that radiates

equally in all directions and has no losses [13]. The effective isotropic radiated power (EIRP) is calculated as

$$EIRP = \frac{P_T G_T}{L_T} = P_{TI} \quad (2.10)$$

where P_{TI} is the effective isotropic transmit power. Similarly, the effective isotropic received power is P_{RI} , where

$$P_{RI} = \frac{P_R L_R}{G_R}. \quad (2.11)$$

The path loss can be defined as the ratio between the transmitted and the received EIRP. By defining as a ratio, path loss can be expressed independent of system parameters. The loss experienced in an idealized system, where the feeder losses are zero and the antennas are isotropic radiators ($G_{T,R} = 1$, $L_{T,R} = 1$), can be defined as

$$L = \frac{P_{TI}}{P_{RI}} = \frac{P_T G_T G_R}{P_R L_T L_R}. \quad (2.12)$$

The propagation loss, thus defined, can be used to describe the propagation medium essentially independently of the system gains and losses. The main goal of path loss modeling is to predict L as accurately as possible. Path loss is one the key inputs of location estimation methods. The path loss in decibels can be expressed as

$$L_{dB} = 10 \log \left(\frac{P_{TI}}{P_{RI}} \right). \quad (2.13)$$

2.1.2. Attenuation

Attenuation is the reduction in power density of an electromagnetic wave as it propagates through space. The particles in the troposphere have a wide range of sizes and characteristics. Attenuation is an important consideration in the modern world of wireless telecommunication. Attenuation decreases the range of radio signals. The materials a signal must travel through such as air, wood, concrete, rain etc. affect the distance a signal can reach [14].

2.1.2.1. Free Space Loss. In telecommunications, free-space path loss (FSPL) is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through free space, with no obstacles nearby to cause reflection or diffraction. FSPL does not consider the gain of the antennas used at the transmitter and receiver. Free space path loss is expressed as

$$FSPL = \frac{P_t G_a G_b}{P_r}. \quad (2.14)$$

FSPL is proportional to the square of the distance between the transmitter and receiver, and also proportional to the square of the frequency of the radio signal [13]. By combining equation equations (2.8) and (2.14), FSPL can be expressed as

$$FSPL = \left(\frac{4\pi r}{\lambda} \right)^2 = \left(\frac{4\pi f}{c} \right)^2. \quad (2.15)$$

Expressing FSPL in decibels, with frequency in megahertz and distance R in kilometers, we obtain

$$FSPL = 32.4 + 20 \log R + 20 \log f_{MHz}. \quad (2.16)$$

2.1.3. Absorption

The reduction in power density due to non-free space propagation is called absorption. In cellular networks, the propagation medium is mostly the air though it penetrates through glass, wood, or walls occasionally. For every unit of distance, the signal loses a certain proportion of its remaining energy. The power density flow depends on the distance d , signal frequency, and the properties of the medium γ . Absorption causes the power density flow to decrease proportional to γ^{-d} . Consequently, the loss decibel scale is linear with respect to the distance [14].

In the upper microwave region, such as frequencies above 10 GHz, absorption loss is particularly significant. At long distances in heavy rain conditions, absorption due to atmospheric conditions becomes comparable to the free-space attenuation at these frequencies. Since the frequencies used in telecommunication systems is below 10 GHz, atmospheric absorption is negligible at distances up to 10 km [14].

Absorption may become more effective when the signal passes through other media. In addition, obstructions cause the wave to be reflected, which decreases the amount of energy passing through. When both reflection and absorption are considered, the total attenuation per obstruction is typically 1-20 dB below 10 GHz, and 1-60 dB above 10 GHz.

2.1.4. Diffraction

According to Huygen's principle, all points on a wavefront are point sources of secondary waves propagating to all directions. When a radio wave passes across an edge such as the corner of a building, it bends around the edge and continues to propagate into the area shadowed by the edge. This effect is called diffraction. In Figure 2.3, the radio wave is transmitted from near an obstacle. The wave is transmitted in all directions and the one that passes through the corner become the source of secondary waves. The circles around the corner indicate the region where the secondary waves are transmitted. The corner is only one of the infinite number of such secondary sources on the wavefront [14].

The waves lose more energy when they bend around a corner. Consequently, the areas with corners gain less additional field strength compared to the areas that have no corners. The field strength of secondary waves is much smaller than the one of the primary waves. If there is line of sight between the transmitter and the receiver, the diffracted waves become negligible.

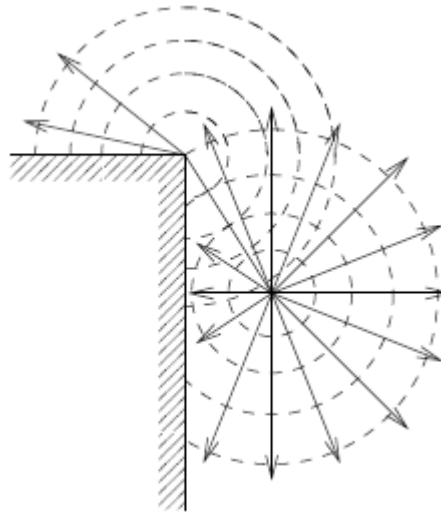


Figure 2.3. Diffraction

2.1.5. Reflection

When a wave meets an obstacle with size much bigger than the wave length, the wave reflects from the obstacle. This effect is called reflection. The wave reflected from the obstacle loses some of its energy according to the properties of the obstacle. If there is line of sight between the transmitter and the receiver, the receiver takes both the incident ray and the reflected ray. In telecommunication systems, the waves usually reflect from the ground. 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 [13, 14].

If there is line of sight, the receiver takes the direct line-of-sight ray and the reflected ray. In the worst case, they cancel each other because of the different phases of the waves. The magnitude of the reflected signal depends on the Fresnel reflection coefficient, which

depends on the properties of the reflecting surface, the frequency of the wave, and the angle of incidence. Roughness is one of the properties of the reflecting surface. Propagating waves scatter in all directions if the reflecting ground is rough. Consequently, the reflection coefficient of a rough surface is smaller than a flat surface. In general, the reflection coefficient is different for the vertical and the horizontal components of the wave. In such cases, reflection can change wave polarization.

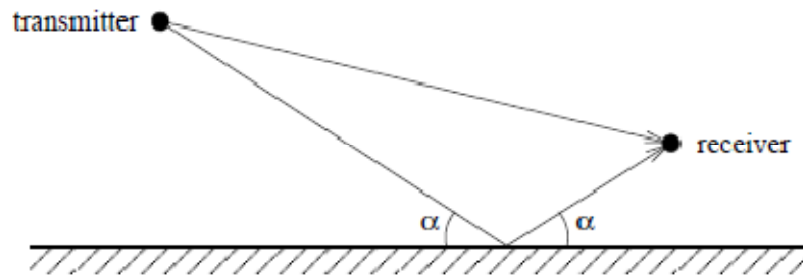


Figure 2.4. Two-ray ground reflection model

2.1.6. Refraction

Electromagnetic refraction is the change in the direction of a ray as it passes obliquely from one medium to another with different velocities of propagation. The velocity of propagating electromagnetic wave is inversely proportional to the density of the medium in which it is propagating. Therefore, when a radio wave passes from one medium into another medium of different density, wave refraction occurs [13].

Refraction is described by Snell's law, which states that the angle of incidence is related to the angle of refraction by

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \quad (2.17)$$

where v_1 and v_2 are the wave velocities through the respective media, θ_1 and θ_2 are the angles between the normal plane and the incident waves respectively, and n_1 and n_2 are the refractive indices.

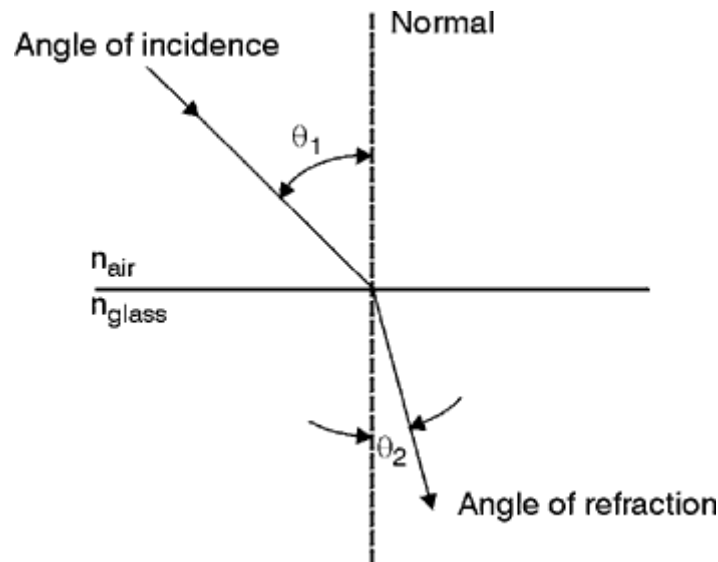


Figure 2.5. Refraction

2.2. Propagation Models

Radio wave propagation models are useful in cell planning, bandwidth allocation, and location estimation. Propagation models are used to predict the behavior of radio waves, their paths and the received signal strength. It is also possible to predict polarization, time dispersion, frequency selectivity and other properties that affect the performance of communication systems. The propagation models are also applicable to telecommunication, broadcasting, private mobile, radio and fixed wireless access applications including WiMax.

Various propagation models are proposed in the literature. These models can be grouped into two categories as deterministic and stochastic models. The deterministic models give a deterministic description of the environment where the signal propagates. The stochastic models give the statistical model of the environment. This section

introduces propagation methods for predicting the path loss encountered in cellular networks.

The field strength is described as a function of the distance between the transmitter and the receiver in propagation models. In addition, propagation models usually include some factors such as ground properties and reflection coefficients. Okumura's model is one of the propagation models that depends on the distance between the transmitter and the receiver. It uses free space loss and correction factors which are determined by experiments. Okumura describes the correction factor as a function of the distance and the frequency of the signal. Okumura proposed different graphs for different environments such as urban, suburban and open areas. Then, Hata analyzed the empirical data of Okumura and proposed formulas for different environment types which closely fit Okumura's graphs. Hata model was proposed for frequencies below 200MHz and above 400MHz. With the Cost 231 model, Hata model is extended to cover the frequencies between 1500MHz and 2000MHz.

2.2.1. The Okumura-Hata Model

Okumura performed a series of measurements in and around Tokyo city between 200 MHz and 2 GHz. Okumura's model is a fully empirical prediction method. There is no mathematical formula proposed for its predictions. The predictions are not based on a physical model such as the plane earth loss. Okumura proposes a series of graphs to visualize his predictions. Hata approximated the graphs of Okumura and proposed the formulas which are currently used in the literature. The combination of these two works has made them the most widely used prediction model. It is often used as a standard against which new approaches are evaluated [13, 44].

The method divides the prediction area into a series of environment types, namely open, suburban, and urban. Open area is described as an open space, no tall trees or buildings in path, plot of land cleared for 300–400 m ahead such as farmland, rice fields, and open fields. Suburban area is described as a village or highway scattered with trees and houses, some obstacles near the mobile but not very congested. Urban area is described as

a built up city or large town with large buildings and houses with two or more floors, or larger villages with close houses and tall, thickly grown trees [11].

Okumura's predictions of median path loss (in decibels) are usually calculated using Hata's approximations as follows for urban areas

$$L_{urban} = 69.55 + 26.16 \log(f_c) - 13.82 \log(h_b) - a(h_m) + (49.9 - 6.55 \log(h_b)) \log R \quad (2.18)$$

and the correction factor for vehicular station antenna height for medium-small cities

$$a(h_m) = (1.1 \log f_c - 0.7) h_m - (1.56 \log f_c - 0.8) \quad (2.19)$$

for large cities

$$a(h_m) = 8.29 (\log 1.54 h_m)^2 - 1.1 \quad : \quad f_c \leq 200 \text{ MHz}, \quad (2.20)$$

$$a(h_m) = 3.2 (\log 11.75 h_m)^2 - 4.97 \quad : \quad f_c \geq 400 \text{ MHz}. \quad (2.21)$$

Okumura's path loss for suburban areas is calculated as

$$L_{suburban} = L_{urban} + (2 \log(f_c / 28))^2 - 5.4 \quad (2.22)$$

and for open areas

$$L_{open} = L_{urban} - 4.78 (\log f_c)^2 + 18.33 \log f_c - 40.94 \quad (2.23)$$

where $150 \text{ MHz} \leq f_c \leq 1500 \text{ MHz}$, $30 \text{ m} \leq h_b \leq 200 \text{ m}$, $1 \text{ m} \leq h_m \leq 10 \text{ m}$, and $R > 1 \text{ km}$ and

- L – path loss (dB),

- f_c – operating frequency (MHz),
- $a(h_m)$ – correction factor for mobile unit antenna height (dB),
- R – distance from transmitting station (m),
- h_b – transmitting station antenna height (m),
- h_m – mobile unit antenna height (m).

In this model, base station antenna height, h_b , is defined as the height above the average ground level in the range 3–10 km from the base station. Therefore, the direction of the mobile from the base may affect h_b . The height gain factor varies between 6 dB and 9 dB as the height increases from 30 m to 1 km. Also, this factor may change upon the range according to measurements [13].

Okumura found that mobile antenna height gain is 3 dB up to h_m 3.25 m and 8 dB beyond. It depends partially upon urban density, apparently as a result of the effect of building heights on the angle-of-arrival of wave energy at the mobile and the consequent shadow loss variation. Urban areas are therefore subdivided into large cities and medium/small cities, where an area having an average building height in excess of 15 m is defined as a large city [13].

Since the measurements of Okumura were done in and around Tokyo city, this model contains the characteristics of the city. Consequently, the success of Okumura's predictions may change according to city which the formula is applied. To overcome this city characterization problem, other researchers tried to extend Okumura model. Kozono proposed including a measure of building density to Okumura's model [16] but such approaches have not gained common acceptance.

The Okumura–Hata model with the related corrections is probably the single most common model used in designing real systems. Several commercial prediction tools use the variations of this model and optimize for the particular environments they are catering for. They form a basis for their predictions by modifying and optimizing Okumura's model.

2.2.2. The Cost 231–Hata Model

The Okumura–Hata model for medium to small cities has been extended to cover the band $1500 \text{ MHz} < f_c < 2000 \text{ MHz}$ [17].

$$L_{urban} = 46.3 + 33.9 \log(f_c) - 13.82 \log(h_b) - a(h_m) + (49.9 - 6.55 \log(h_b)) \log R + C \quad (2.24)$$

$$C = \begin{cases} 0 \text{ dB for medium cities and suburban areas} \\ 3 \text{ dB for metropolitan areas} \end{cases} \quad (2.25)$$

2.2.3. The Walfisch–Bertoni Model

The Walfisch-Bertoni model is one of the popular propagation models. LOS and NLOS situations are considered in this model in a frequency range of 800 to 2000 MHz. The Walfisch–Bertoni model is the first model to have accounted for the effects of buildings on propagation in urban areas [18]. The model is expressed as

$$L_{ex} = 57.1 + A + \log f_c + 18 \log R - 18 \log(h_b - h_0) - 18 \log \left[1 - \frac{R^2}{17(h_b - h_0)} \right] \quad (2.26)$$

where

$$A = 5 \log \left[\left(\frac{d}{2} \right)^2 + (h - h_m)^2 \right] - 9 \log d + 20 \log \{ \tan^{-1} [2(h - h_m)/d] \}. \quad (2.27)$$

The parameters and the model are shown in Figure 2.6.

Since the Walfisch-Bertoni model shows the effect of buildings, it identifies the important physical properties that effect the propagation. One drawback of this model is that it is designed for urban areas where detailed information such as street widths,

building heights, etc. is needed. The requirement of extra information makes this model less applicable with respect to Okumura's model.

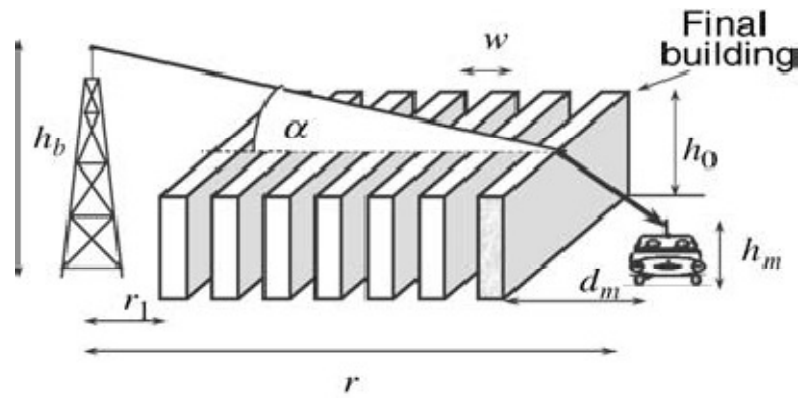


Figure 2.6. The Walfisch-Bertoni propagation model

3. LOCATION ESTIMATION METHODS

Location estimation is defined as the method for obtaining the location of MS. There are several location estimation methods in the literature [1]. Majority of location estimation applications utilize the GPS which provides location estimates with an accuracy of a couple of meters. Alternatives to satellite-based systems are developed to avoid problems such as lack of coverage between high buildings and indoors, and high energy consumption of devices. These techniques use signals between the MS and BTS. The location estimation methods are categorized as network-based, handset-based, and hybrid methods.

Network-based methods use the operator's network infrastructure to find the location of MS. The measurements are done by one or more BTSs and sent to the location server where the position of the MS is estimated. By using the location of the serving base station of a cellular network, a very rough location estimate can be obtained easily. More accurate estimates require measuring the other properties of the signals transmitted between the MS and BTS. The advantage of network-based methods is that they can be implemented without changing the handsets. As a result, it is possible to locate all MSs under the network. The accuracy of network-based methods varies according to the used algorithms and the infrastructure of the operator.

Handset-based methods use the capabilities of the handset to locate the MS. These methods give more accurate results than network-based methods. These methods determine the location of the MS by computing its location by cell identification, signal strengths of the home and neighboring cells, or the latitude and longitude. They require some specific features to work in legacy handsets. Since it requires a change in the handsets, the adoption of handset-based methods may be longer due to the slow spread of new handsets.

Hybrid-based methods use a combination of network-based and handset-based methods to locate the MS. Assisted-GPS is one such method. It uses both GPS and network information to compute the location. Hybrid-based methods give the better

accuracy compared to prior methods, but they have the same drawbacks as the network-based and handset-based methods.

Location estimation methods can also be categorized according to their measurement principles: multilateral, unilateral, and bilateral. Multilateral methods are network-based methods in which several BTSs make measurements together. Unilateral methods may be handset-based or hybrid methods in which MS takes measurements from serving and neighbor BTSs. Finally, if MS takes measurements from a single BTS or a single BTS takes measurements from an MS, it is called a bilateral method.

3.1. Cell Global Identification and Timing Advance

The method which uses CGI and TA together is one of the simplest methods in terms wide applicability. This bilateral method can be implemented either as network-based or handset-based. CGI of the serving cell of MS is used to locate the MS [42]. It is often complemented with the timing advance information to increase the accuracy.

TA is the measured time that passes while a signal propagates from the BTS to MS. TA parameter is accepted as an estimate of the distance from the MS to the serving BTS. TA values are divided into 64 slots and can take values from 0 to 63. Each increase in TA value corresponds to moving away from serving BTS by 550m. With the use of TA value, the location of the MS can be narrowed to a circle. This information is already in the network and the accuracy increases when the cells become smaller. The accuracy is dependent on the cell size and varies from 10m, a micro cell in a building, to 500m, in a large outdoors macro cell. In rural areas, it may be up to 35km.

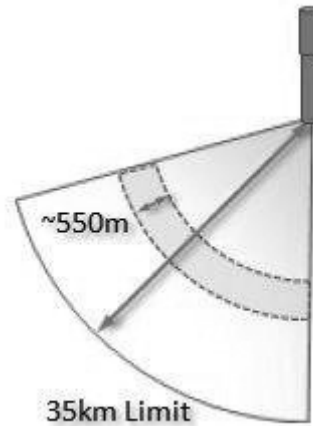


Figure 3.1. CGI and Timing Advance Method

3.2. Time of Arrival

Time of arrival method can be implemented both as a unilateral mobile-assisted method or a multilateral network-based method. TOA uses the one way trip time of a signal between the BTS and MS. In unilateral mobile-assisted implementation, MS takes a signal from BTS and it records the time of arrival of the signal. Then, it sends this data to the location server. In multilateral network-based implementation, MS sends a burst signal to the BTS, and BTS records the time of arrival then send this data to the location server.

TOA method requires values from at least three BTSs. By measuring the time of arrival of a signal, the distance between the MS and BTS can be calculated by simply multiplying the time with the speed of light. After calculating the distances from BTSs, the MS can be located by using the least squares method [19, 20].

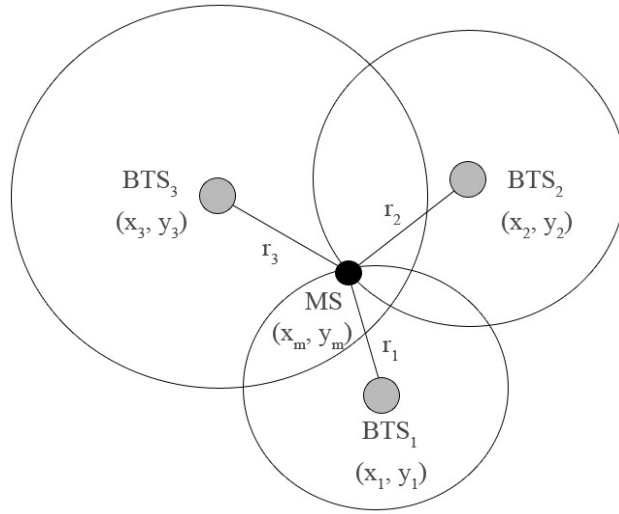


Figure 3.2. TOA method

The distance r_i between the MS and the BTS, can be calculated as

$$r_i = (t_i - t_0)c \quad (3.1)$$

where c is the speed of light, t_0 is the time when the transmission started and t_i is the time of arrival of the received signal. The coordinates (x_m, y_m) of the mobile user can be found by using the following distance equations:

$$r_1 = x_m^2 + y_m^2 \quad (3.2)$$

$$r_2 = (x_2 - x_m)^2 + (y_2 - y_m)^2 \quad (3.3)$$

$$r_3 = (x_3 - x_m)^2 + (y_3 - y_m)^2. \quad (3.4)$$

Subtracting Equation (3.2) from Eq (3.3) gives

$$r_2^2 - r_1^2 = x_2^2 - 2x_2x_m + y_2^2 - 2y_2y_m. \quad (3.5)$$

Similarly, subtracting Equation (3.15) from Eq (3.17) gives

$$r_3^2 - r_1^2 = x_3^2 - 2x_3x_m + y_3^2 - 2y_3y_m. \quad (3.6)$$

The above two equations can be written in matrix form as

$$\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \end{bmatrix} = \frac{1}{2} \begin{bmatrix} K_2^2 - r_2^2 + r_1^2 \\ K_3^2 - r_3^2 + r_1^2 \end{bmatrix} \quad (3.7)$$

where

$$K_i^2 = x_i^2 + y_i^2. \quad (3.8)$$

Then, Equation (3.7) can be rewritten as

$$\mathbf{H}\mathbf{x} = \mathbf{b} \quad (3.9)$$

where

$$H = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix}, \quad x = \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \quad b = \begin{bmatrix} K_2^2 - r_2^2 + r_1^2 \\ K_3^2 - r_3^2 + r_1^2 \end{bmatrix}. \quad (3.10)$$

Thus, the least-squares solution is

$$\mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{b}. \quad (3.11)$$

TOA method requires synchronization between the clocks of the BTS and MS since the distance between BTS and MS is expressed as multiplication of the difference with the speed of light. Even microseconds may result in an error in the location estimation of the TOA method. Even the MS clock itself might have a drift of microseconds which makes TOA method inappropriate. TDOA overcomes this problem by using the time differences, and it doesn't require the synchronization of the clocks of BTS and MS.

3.3. Time Difference of Arrival

TDOA method uses the difference values of the time between BTS and MS. By using the time difference values, it eliminates the errors that may occur due to the synchronization problems between the BTS and MS. TDOA values of two BTSs form hyperbolae as shown in Figure 3.3 [20].

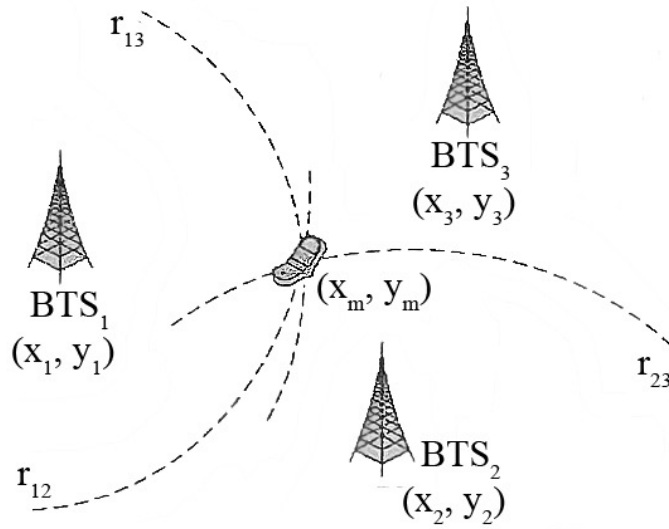


Figure 3.3. TDOA method

The hyperbolic curves r_{ij} are defined as:

$$r_{ij} = \sqrt{(x_i - x_m)^2 - (y_i - y_m)^2} - \sqrt{(x_j - x_m)^2 - (y_j - y_m)^2} \quad (3.12)$$

where r_{ij} is the difference in the time delays between the MS and BTSs i and j . The time differences between the TOAs of the MS signal at BS_i and BS_1 is

$$t_i - t_1. \quad (3.13)$$

The distance differences are

$$r_{i1} = r_i - r_1 = (t_i - t_0)c - (t_1 - t_0)c = (t_i - t_1)c \quad (3.14)$$

Since TDOA uses the time differences, t_0 values cancel each other as seen from equation (3.14). Consequently, the distance differences are not dependent on t_0 value, and it doesn't require clock synchronization between BTS and MS. Rewriting the equation (3.3) in terms TDOA measurement r_{21} :

$$(r_{21} + r_1)^2 = K_2^2 - 2x_2x_m - 2y_2y_m + r_1^2. \quad (3.15)$$

Rearranging gives

$$-x_2x_m - y_2y_m = r_{21}r_1 + \frac{1}{2}(r_{21}^2 - K_2^2). \quad (3.16)$$

Rewriting the equation (3.4) in terms of TDOA measurements r_{31} , we get

$$-x_3x_m - y_3y_m = r_{31}r_1 + \frac{1}{2}(r_{31}^2 - K_3^2). \quad (3.17)$$

Rewriting the equations (3.16) and (3.17) in the matrix form gives

$$Hx = r_1c + d \quad (3.18)$$

where

$$c = \begin{bmatrix} -r_{21} \\ -r_{31} \end{bmatrix}, d = \frac{1}{2} \begin{bmatrix} K_2^2 - K_{21}^2 \\ K_3^2 - K_{31}^2 \end{bmatrix}. \quad (3.19)$$

Then, the solution of equation (3.18) for x is expressed as

$$x = r_1 H^{-1} c + H^{-1} d. \quad (3.20)$$

The final solution for x is achieved by finding the r_1 value by substituting equation (3.20) into equation (3.2). Equation (3.18) still holds in the existence of more than three BTSs with

$$H = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \\ \dots & \dots \end{bmatrix}, c = \begin{bmatrix} -r_{21} \\ -r_{31} \\ -r_{41} \\ \dots \end{bmatrix}, d = \frac{1}{2} \begin{bmatrix} K_2^2 - K_{21}^2 \\ K_3^2 - K_{31}^2 \\ K_4^2 - K_{41}^2 \\ \dots \end{bmatrix}, \quad (3.21)$$

similarly yielding

$$x = (H^T H)^{-1} H^T (r_1 c + d). \quad (3.22)$$

3.4. Angle of Arrival

Angle of arrival method is a multilateral and network-based method since several BTSs take measurements coming from the MS. AOA method requires extra hardware such as antenna arrays on the network side for angle measurements. AOA method estimates the location of an MS by using the AOA information of the signal from two BTSs [20].

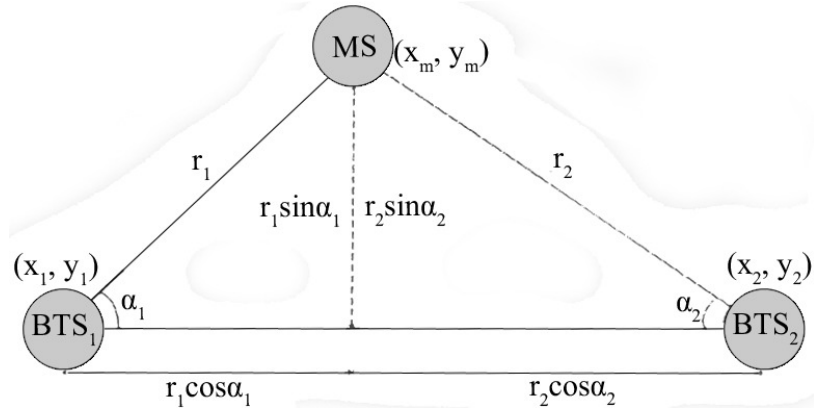


Figure 3.4. AOA method

By combining the knowledge of AOA from two BTSs as shown in Figure 3.4, the location of the MS can be calculated as follows. From the Figure 3.4, we can write

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} r_1 \cos \alpha_1 \\ r_1 \sin \alpha_1 \end{bmatrix} \quad (3.23)$$

and

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} r_2 \cos \alpha_2 \\ r_2 \sin \alpha_2 \end{bmatrix}. \quad (3.24)$$

Similarly, if we generalize the equation (3.24) for any BTS

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} r_i \cos \alpha_i \\ r_i \sin \alpha_i \end{bmatrix} \quad (3.25)$$

These equations can be written as

$$\mathbf{H}\mathbf{x} = \mathbf{b} \quad (3.26)$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \dots \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \mathbf{b} = \begin{bmatrix} r_1 \cos \alpha_1 \\ r_2 \cos \alpha_2 \\ x_2 + r_2 \cos \alpha_2 \\ y_2 + r_2 \cos \alpha_2 \\ \dots \\ x_n + r_n \cos \alpha_n \\ y_n + r_n \cos \alpha_n \end{bmatrix}. \quad (3.27)$$

Then, the least squares solution for $\mathbf{x}$ is

$$\mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{b}. \quad (3.28)$$

AOA method gives accurate results if there is LOS condition. If there is no LOS condition, then the accuracy of AOA method highly decreases. Since dense urban areas contain obstacles, the performance of AOA method is worse than the suburban and rural areas. The accuracy is highly dependent to the AOA information so the calibration of antenna arrays is very important. Such a small error in the calibration may result serious accuracy losses.

3.5. Assisted GPS

Assisted GPS is a hybrid method that requires a GPS receiver to be integrated into the handsets. On the network side, at least an assistance server is needed to locate the MS. The integrated GPS receiver performs the tasks required to make range measurements and position solutions. The mission of the assistance server is to help the MS in its calculations. Assistance server is able to communicate with the GPS receiver on the handset on a wireless link. Assistance server shares the tasks of the handset, and the calculations can be made quickly and efficiently than the unassisted case. Figure 3.5 shows the A-GPS architecture [23].

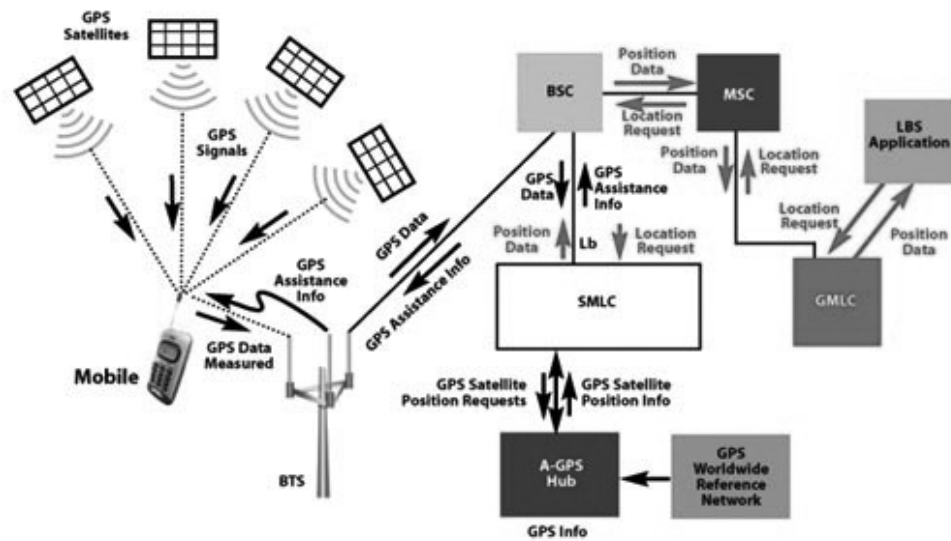


Figure 3.5. A-GPS architecture

A-GPS performs better in terms of processing time and accuracy than a stand-alone mode GPS since assistance server helps the MS. Assistance server may provide three types of data:

- Initial position,
- Time estimate,
- For A-GPS only receivers, satellite selection, range, and range-rate information.

Additionally, assistance server can compute the position solutions. Thus, the only thing that the GPS receiver should do is to collect range measurements.

Mobile handsets must have two antennas when they have GPS receivers. GPS uses higher frequency levels than GSM. Since GPS receivers consume relatively high battery, the handsets that have GPS receivers must have higher batter capacities than the legacy handsets.

3.6. Signal Strength

Signal strength method is a unilateral method that can be implemented as a hybrid or a handset-based method. Handset-based method implementation requires that BTS coordinates are transmitted to the MS. In active mode, MS continuously transmits measurement reports back to the network. By using these measurement reports, signal strength method can be implemented easily. Since there is no need to change the handsets, it is sometimes called a network-based method although MS performs the measurements. An extension to this system may be modifying MS such that it also sends the measurements in the idle mode.

The distances between the MS and the BTSs can be estimated by using the signal strength measurements from several BTSs. If the signal levels from three different BTSs are known, the location of the MS can be calculated by finding the unique intersection point of the three circles. In an ideal environment, by considering free space loss, the location of the MS can be calculated. Practically, there are other factors that affect the signal strength values. Consequently, an environment-dependent propagation model should be used for the relation of the received signal strength and the distance between the BTS and the MS. In urban areas, the received signal strength generally decreases more rapidly with distance than in suburban or rural areas [19].

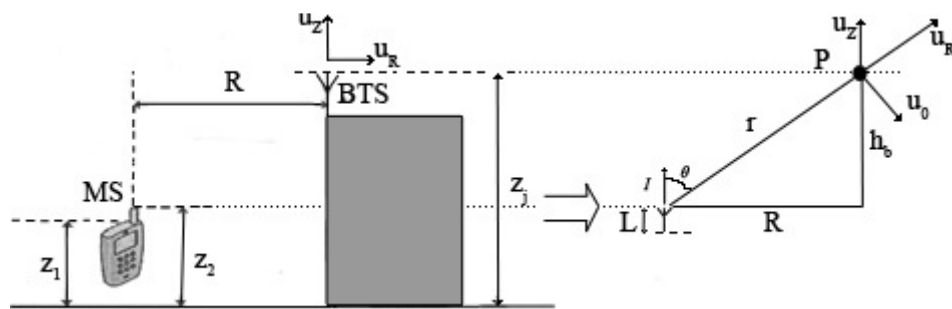


Figure 3.6. MS-BTS geometry

The geometry of the MS and BTS is given in Figure 3.6. The electric field radiated by the antenna of the MS can be written as [19]:

$$E(r, \theta) = j \frac{60I_0}{r} \left[\frac{\cos\left(\frac{kL}{2} \cos \theta\right) - \cos\left(\frac{kL}{2}\right)}{\sin \theta} \right] e^{-jkr} \quad (3.29)$$

The current can be expressed by $I_0 \cos(\omega t - kr)$ where I_0 is the amplitude and ω is the frequency. Therefore, the electric field strength can be expressed in the form

$$|E(r, \theta)| = \frac{E_0}{r} \left[\frac{\cos\left(\frac{kL}{2} \cos \theta\right) - \cos\left(\frac{kL}{2}\right)}{\sin \theta} \right] \quad (3.30)$$

$$|E(R)| = E_0 \left[\frac{\cos\left(\frac{kL}{2} \frac{h}{\sqrt{R^2 + h^2}}\right) - \cos\left(\frac{kL}{2}\right)}{R} \right] \quad (3.31)$$

where

$$E_0 60I_0, I_0 = \sqrt{\frac{P}{R_s}}. \quad (3.32)$$

Here, P is the signal power radiated by the MS and R_s is the radiation resistance of the MS antenna.

For the locating 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 at BTS for MS are

$$E = [E_1, E_2, E_3] \quad (3.33)$$

$$E_m = [E_{m1}, E_{m2}, E_{m3}] = \alpha E + n \quad (3.34)$$

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

$$R = [R_1, R_2, R_3] \quad (3.35)$$

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

$$E(R) = E(R_0) + \frac{\partial E}{\partial R} \Delta R + O(\Delta R) = E_0 + J_0 \Delta R + O(\Delta R) \quad (3.36)$$

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

$$J_0 = \begin{bmatrix} \frac{\partial E_1}{\partial R_1} & 0 \\ \frac{\partial E_2}{\partial R_2} & \\ 0 & \frac{\partial E_3}{\partial R_3} \end{bmatrix}. \quad (3.37)$$

From equation (3.31) we have

$$\frac{\partial E}{\partial R} = E_0 \left[\frac{kLh}{2(R^2 + h^2)^{3/2}} \sin\left(\frac{kLh}{2\sqrt{R^2 + h^2}}\right) - \frac{1}{R^2} \left(\cos\left(\frac{kLh}{2\sqrt{R^2 + h^2}}\right) - \cos\left(\frac{kL}{2}\right) \right) \right]. \quad (3.38)$$

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

$$\varepsilon = E_m - E(R) = E_m - E_0 - J_0 \Delta R. \quad (3.39)$$

The minimum of ε^2 is obtained when

$$\frac{\partial u}{\partial R} = 0. \quad (3.40)$$

The main iteration equation for R is obtained as

$$R^{k+1} = R^k + (J_k^T J_k)^{-1} J_k^T (E_m - E^k). \quad (3.41)$$

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

3.7. Fingerprinting

Fingerprinting based methods appear in the literature also under the names “database correlation” and “pattern matching.” Fingerprint methods have been preferred as they perform better in areas with severe multipath propagation compared to others [36,37,38,39,43]. Fingerprinting method can be applied to any cellular network. In fingerprinting, there is a location database which is responsible for storing the signal information sent by the MS. This database should contain the signal information by not losing any information that can be used in location estimation. The signal information sent by MS is called the fingerprint. Fingerprints must contain any data that can be measured by the MS. Measurements performed by the network may also be used [24].

MS sends the necessary measurements that are performed and transmits them to the location server. Then, the location server calculates the MS location by comparing the

transmitted fingerprint with the fingerprints in 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. Distributed processing may solve any performance problems. Also, some area based algorithms may decrease the number of comparisons and the computing power requirement.

In network planning, similar information that is stored in the fingerprint database is also needed. Therefore, the creation and maintenance of the database also support network planning. Any cellular network that has a network planning method can use fingerprinting method with its data. Fingerprints for the database can be collected either by measurements or by a computational network planning tool. Taking substantial amount of measurements is hard and time consuming, but it produces more accurate fingerprint data compared to the network planning data. Also, a combination of measured and computed fingerprints can be used. For the accuracy of the fingerprinting method, the database must be up-to-date. The changes in the network affect the accuracy of fingerprinting.

4. FINGERPRINTING METHODS

The key idea of fingerprinting is to store the location-dependent network parameters seen by the MS, for the whole coverage area of the location system, in a database that is used by a location estimation algorithm. These stored signal information samples in the database should contain any location dependent network parameter [26]. Fingerprinting consists of two parts. The first part is finding the points which have the most similar RSS characteristics with the MS measurements. Fingerprint matching algorithms are used in the first part to find the points. The second part is the finding the location of the MS by using the points found by the first part. Location estimation algorithms are used in the second part.

When the MS needs to be located, the necessary measurements are performed and transmitted to the location server. The location server compares the received information with the fingerprints in the database and calculates the MS location. Hence, the fingerprint matching algorithm and the location estimation algorithm play a major role in the accuracy of the estimation.

4.1. Fingerprint Matching Methods

After the collection of the fingerprints, a fingerprint database is created. When the MS needs to be located, fingerprint matching technique or a database correlation method could be applied to find the difference between the RSS measurements at the MS to be located (observation) and each fingerprint in the database. By using this difference, the nearest fingerprints for the MS location could be identified and the location could be estimated using these nearest neighbors. This section presents five types of deterministic fingerprint matching algorithms.

4.1.1. GSM for Location Method

This method is a simple correlation-type algorithm for finding the best match between the fingerprint to be located and the fingerprints of the database. A difference

between the fingerprint to be located and each fingerprint of the database is calculated as [26]

$$d(k) = \sum_i (f_i - g_i(k))^2 + p(k) \quad (4.1)$$

where f_i is the signal strength of the request fingerprinted on the i^{th} broadcast control channel, $g_i(k)$ is the signal strength of k^{th} fingerprint in the database on the same channel, and the summation is taken over channels that are found in both fingerprints. Each channel that is found in only one of the fingerprints contributes to the penalty term, $p(k)$. The coordinates of the fingerprint in the database that minimizes this difference are returned as the resulting location. It should be noted that the database search can be limited, based on LAC and cell ID, to a relatively small area. Also other types of algorithms, e.g. pattern matching, could be used.

4.1.2. Euclidian and Manhattan Distance Method

Euclidian and Manhattan distance methods are based on measuring the signal distance between each fingerprint in the database and the observation. The general equation to calculate the signal distance between the observation and the database fingerprint can be given as

$$d(k) = \left(\sum_{i=1}^n |f_i - g_i(k)|^q \right)^{1/q} \quad (4.2)$$

where f_i is the RSS from the i^{th} accessible cell in the observation, $g_i(k)$ is the RSS from the same cell in the k^{th} database fingerprint [28]. The Manhattan distance and the Euclidean distance are two widely used distance measurement methods corresponding to $q=1$ and $q=2$, respectively.

4.2.1.1. Euclidian Distance. By replacing q with 2 in equation (4.2), we get the Euclidian distance method which is [28]

$$d(k) = \sqrt[2]{\sum_i (f_i - g_i(k))^2 + \sum_j (f_j - I_{\max})^2 + \sum_k (I_{\max} - g_k(k))^2} \quad (4.3)$$

4.2.1.2. Manhattan Distance. By replacing q with 1 in equation (4.3), we get the Manhattan distance method which is [28]

$$d(k) = \sum_i |f_i - g_i(k)| + \sum_j |f_j - I_{\max}| + \sum_k |I_{\max} - g_k(k)| \quad (4.4)$$

4.1.3. DCM for Multi-System Location

Equation (4.1) could be directly applied to calculate the signal distance only if same set of cells appear in both vectors. Cells which appear only in one vector would contribute as penalty terms. Thus, equation (4.1) should be modified to handle these penalty terms. Then, the problem arises on how to handle these penalty terms. If a particular cell does not appear in one vector it is reasonable to think that the RSS from that cell is extremely low in this vector. Therefore, that missing signal strength value could be replaced by some threshold RSS value. Based on this approach the difference between the observation and each fingerprint in the database can be calculated using equation 4.5 [29].

$$d(k) = \sum_{i=1}^n (f_i - g_i(k))^2 + \sum_j (f_j - I_{\max})^2 + \sum_k (I_{\max} - g_k(k))^2 \quad (4.5)$$

where

- $d(k)$ - the difference between the observation and the k^{th} fingerprint in the database.
- f_i - the averaged RSS from the i^{th} accessible cell in the observation

- $g_i(k)$ - the averaged RSS from the same cell in the k^{th} fingerprint in the database
- f_j - the averaged RSS from the j^{th} accessible cell in the observation which is not accessible in the k^{th} fingerprint in the database
- $g_k(k)$ - the averaged RSS from the k^{th} accessible cell in the fingerprint in the database which is not accessible in the observation
- $I_{\max}$ - replacement for missing signal strength value by a threshold RSS value

4.1.4. Novel Cost Function for DCM

Since several measurements are obtained in each test location and several measurements are used to create a fingerprint, it is possible to modify the penalty term by applying a weight based on the contribution of the corresponding cell in the observation or in the fingerprint. This concept gives rise to a novel cost function as [27].

$$d(k) = \sum_i |f_i - g_i(k)| + \sum_j |(f_j - I_{\max})| \times w_j + \sum_k |(I_{\max} - g_k(k))| \times w_k \quad (4.6)$$

where w_j and w_k are corresponding weights of the penalty terms and defined as the contribution of the penalty cell / total number of measurements. The summation is taken over all the cells that are found in the observation and the fingerprint in the database. The database fingerprints with lower $d(k)$ values are considered as the nearest neighbors of the location to be estimated.

4.1.5. DCM by Using Average RSS and Standard Deviation

This is based on defining a valid RSS range for each cell in each fingerprint using the averaged RSS and the standard deviation of the RSS of the corresponding cell. The standard deviation of the RSS of each accessible cell in each fingerprint is to be calculated and stored in the database. Then, it is possible to determine whether an input RSS value lies within this valid RSS range using [30]

$$c(k) = \sum_i \delta(f_i, g_i(k), std_i(k)) \quad (4.7)$$

where

- $c(k)$ - the total number of cells in the observation which falls within the valid RSS of corresponding cells in the k^{th} fingerprint in the database.
- f_i - the averaged RSS from the i^{th} accessible cell in the observation
- $g_i(k)$ - the averaged RSS from the same cell in the k^{th} fingerprint in the database
- $std_i(k)$ - the standard deviation of the same cell in the k^{th} fingerprint in the database
- $\delta(x, y, z)$ - a function which returns 1, if x falls within the region $y-z$ and $y+z$, otherwise it returns 0.

The summation is taken over the cells that are common in the observation and the fingerprint in the database. The fingerprints in the database with higher $c(k)$ values are considered as the nearest neighbors of the location to be estimated.

4.2. Location Estimation Methods

After identifying the nearest neighbors, there are several algorithms for estimating the location of the MS during the location estimation. The basic one is the Nearest Neighbor (NN) approach. If multiple nearest neighbors are chosen, a K-Nearest Neighbor (kNN) or a Weighted K-Nearest Neighbor (WkNN) method could be applied. Performance of the NN and the WkNN approaches was evaluated during this research.

4.2.1. Nearest Neighbor (NN) method

In the nearest neighbor method, best matching fingerprint in the database is found. The position of the MS is taken as the position of the fingerprint found. Thus, the error distance is calculated as the distance rate between the Ms and the fingerprint. Consequently, in this case the sampling distance rate becomes important. If the measurements are taken in short distances, the error in this method can be decreased.

4.2.2. Weighted k Nearest Neighbor (WkNN) method

Here, k ($k \geq 2$) nearest fingerprints are chosen and the location is estimated using the Weighted k-Nearest neighbors (WkNN) method [28] as

$$p = \sum_i p_i w_i \quad (4.8)$$

where

- p - estimated location
- p_i - location of the i^{th} nearest neighbor
- w_i - weight of the i^{th} nearest neighbor

WkNN method was applied for two neighbors and up to five neighbors

$$2 \leq k \leq 5 \quad (4.9)$$

and the results are calculated by the weight formula

$$w_k = \frac{\left(\frac{1}{d(k)} \right)}{\sum_k \frac{1}{d(k)}} \quad (4.10)$$

5. LOCATION ESTIMATION BY FINGERPRINTING (LEF)

In this thesis, we propose a novel database correlation algorithm name Location Estimation by Fingerprinting (LEF). LEF method consists of four phases. The methods used in these phases are explained in chapter 4.

- Phase one – Data Collection
- Phase two – Fingerprint Filtering
- Phase three – Fingerprint Matching
- Phase four – Location Estimation

Phase one is the data collection phase. To be able to create a fingerprint database, several real measurements are taken from the city of Izmir and Istanbul in Turkey. Phase two is the fingerprint filtering part. Before applying our algorithm we have to filter the database to decrease the processing time of our algorithm and increase the accuracy of the algorithm. Phase three is applying fingerprint matching algorithm. After filtering the fingerprints, the proposed fingerprint matching algorithm is compared with the several best-known algorithms. Phase four is the location estimation by using the nearest points found. Working mechanism of the algorithm is shown in Figure 5.1.

5.1. Data Collection

In phase one, raw measurement data have been gathered using a cell phone and a GPS device connected to a laptop running Ericsson TEMSTM Investigator software as depicted in Figure 5.1. The data are collected in various locations of Istanbul and Izmir with different environmental characteristics.

Izmir has a complex geography for location estimation. There is a bay in the middle of the city so the city is almost divided into two areas and the city also has hills. The data are collected from both sides of the bay, from seashore, inside the city, and also from the rural parts of the city. Istanbul has also a challenging topography for location estimation. The Bosphorus splits the city into two parts with a narrow strait that is 1-3 km wide and

30 km long. The data are collected from both sides of the Bosphorus and several different parts of the city. These collected data are preprocessed to eliminate the erroneous and deficient data such as missing RSS levels or CGIs. GPS measurements are assumed to be error free and taken as the exact mobile positions.

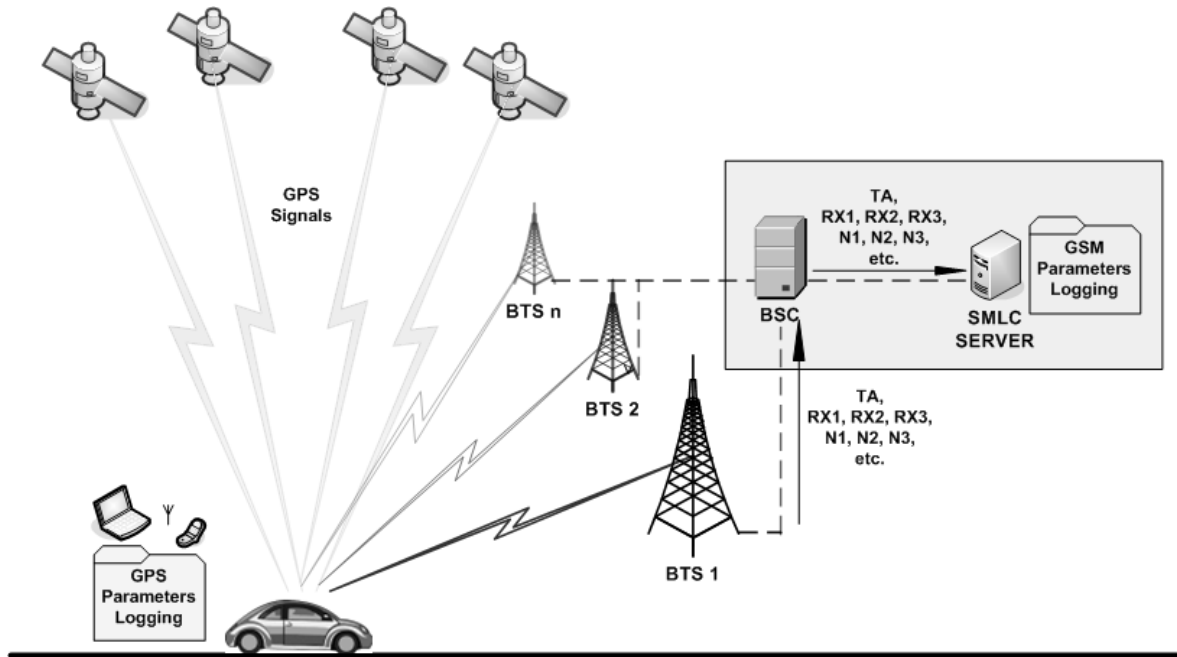


Figure 5.1. Field Measurement Process

Data collection is done as a two-party collaboration process. While making drive-tests and recording the MS coordinates (via GPS) together with time stamps, the SMLC in the GSM network records the other parameters such as TA and Rx values, serving BTS ID, neighboring BTS IDs, etc. sent by the MS together with time stamps. Afterwards, these two sources are fused to get a normalized and proper dataset. In order to eliminate some intrinsic problems in continuous call-based localization scheme, we have switched to the trigger-based scheme. In the 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 on the network side.

The four input dimensions what help us identify the environment at the beginning are as follows:

- TA – Timing advance parameter of the mobile station
- Rx – RSS parameter of the serving cell
- Rx_i – RSS parameter of the i^{th} neighbor station
- $NCELLID_i$ – Neighbour cell id of the i^{th} neighbor station.

5.2. Fingerprint Filtering

In order to reduce the time taken for correlation, the database search could be limited to a small area based on the serving cell of each fingerprint in the database. The reason behind choosing fingerprints that have the same serving cell id is that probably those fingerprints are the closest points to the MS. Firstly, the fingerprints are filtered according to their serving cell ids. Then, a list is created with the filtered fingerprints and the number of fingerprints in the list is checked. In this thesis, we assume that at least one point is enough to use this list in the fingerprint matching algorithm. If there are not enough fingerprints in the filtered list, a new list that consists of the fingerprints that are served by the neighbor cell ids of the MS is prepared. If the conditions are still not met, the algorithm is applied to the whole measurements.

5.3. Fingerprint Matching Algorithm

After filtering the fingerprints in the database, fingerprint matching technique could be applied to find the difference between the Rx measurements at the MS and each fingerprint in the database. Based on this difference, the nearest fingerprints for the MS location could be identified and the location could be estimated using these nearest neighbors [27].

Fingerprint matching algorithms are explained in Chapter 4.1. There are several fingerprint matching algorithms. GSMLocation, ED, MD, and MSL are the frequently applied methods. In this work, these four methods and our proposed method, LEF, are applied to our fingerprint database and their performances are compared in the next chapter.

5.3.1. LEF Fingerprint Matching Algorithm

The fingerprint matching algorithms calculate the correlation between the Rx values of the MS and the fingerprints in the database. The Rx values from the cell that are both measured by the MS and the fingerprint are compared. However, Rx values that exist for only one of them cannot be compared with a value. So, a method should be used include these Rx values into penalty value. Method 3 calculates the difference between Rx values of the MS and the fingerprint as described in equation (4.5) where $I_{\max}$ is the signal threshold.

The threshold value, that gives minimum error in distance value among the tried threshold range, is calculated during analysis and it is observed that the effect of the punishment given by method 3 increases the error in accuracy. To decrease the effect of punishment while still taking it into account, a new fingerprint matching algorithm is proposed. It can be calculated as

$$d(k) = \sum_{i=1}^n (f_i - g_i(k))^2 + \sum_j \left(\frac{C}{f_j} \right)^2 + \sum_k \left(\frac{C}{g_k(k)} \right)^2 \quad (5.1)$$

where C is constant which gives the best results during the experiments in the region.

5.3.2. LEF Training

The punishment values are calculated by dividing a constant, LEF constant, by the received Rx value of the BTS. Since the punishment values affect the correlation between the measurements of MS and the fingerprint, the LEF constant become an important factor in the performance of the LEF algorithm. By running several experiments, it is observed that the LEF constant should resume a value between

$$0 \leq C \leq 5000. \quad (5.2)$$

In order to calculate the optimum LEF constant, the data is randomly divided into five parts. One of these parts is used for training, LEF constant calculation. The algorithm is run for the range of LEF constant and errors in distance for different LEF constants are recorded. After the runs, the LEF constant that gives minimum error in distance among the others is chosen to be used for that area.

5.4. Location Estimation Algorithms

After identifying the nearest neighbors, Nearest Neighbor (NN) approach is applied for estimating the location of the MS. If multiple nearest neighbors are chosen, a Weighted KNearest Neighbor (WkNN) method can be applied [33, 34].

ED, MD, MSL, and LEF are run to find the nearest fingerprint in the database. When these algorithms are run for finding a single fingerprint, the error in distance is calculated by Vincenty's formula [35], an extremely accurate geodesic calculation technique. Then, these algorithms are run to find the k nearest fingerprints in the database where k is

$$2 \leq k \leq 5. \quad (5.3)$$

For calculating the error in distance for these k values, the k-weighted nearest neighbor WkNN formula, equation 4.10, is used. Several experiments are made to find the best k which minimizes the error in distance.

6. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this section, four different methods of fingerprint matching algorithm and our method are analyzed for two different measurement sets. The tests are done for one to five neighbors. The optimum threshold and coefficient values for the measurement sets are calculated and used in experiments. The methods are tested against all the possible values in the range and the ones that gives the minimum error are taken. Finally, the tests are repeated for 10 times and the average of those results is used in the graphs.

6.1. Dataset Analysis

The tests are run for two different datasets. The first dataset is taken from the city of Izmir and the second one is taken from the city Istanbul in Turkey.

6.1.1. Istanbul Dataset

The Istanbul dataset consists of 375 measurements taken from different parts of the city. The region and the fingerprints are shown in Figure 6.1. The TA and Rx histogram of the measurements are shown in Figure 6.2 and 6.3, respectively. To show the effect of number of measurements for fingerprinting, the dataset is divided into several regions according to their serving BTS ids. Table 6.1 shows the number of the measurements for the scenarios of Istanbul dataset.



Figure 6.1. Istanbul measurements

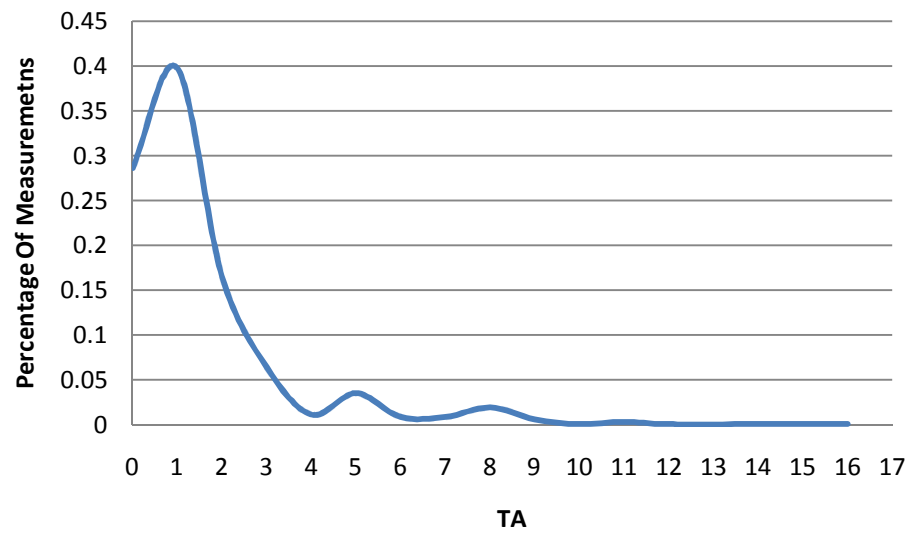


Figure 6.2 TA histogram of Istanbul measurements

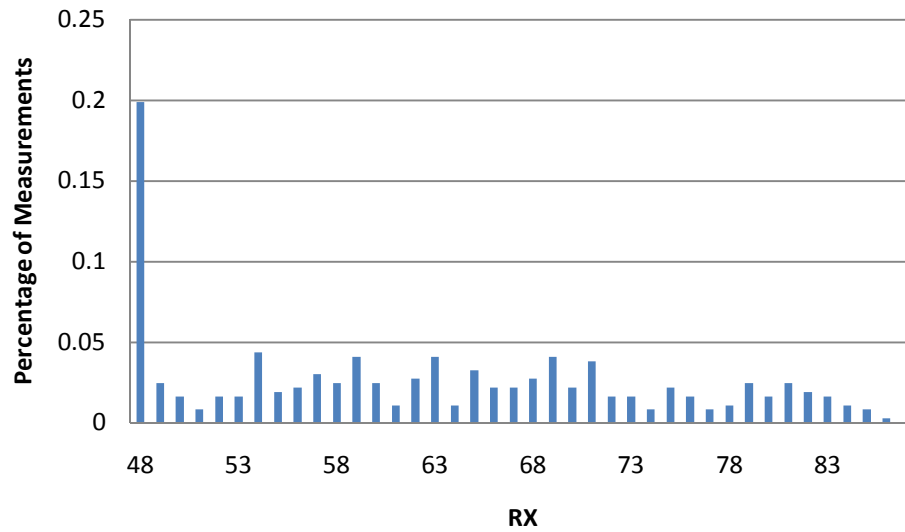


Figure 6.3 Rx histogram of Istanbul measurements

Table 6.1. Number of measurements for same serving BTS in Istanbul

	Number of Measurements
Minimum 1 measurement	375
Minimum 2 measurement	285
Minimum 3 measurement	201
Minimum 4 measurement	156
Minimum 5 measurement	108
Minimum 6 measurement	88
Minimum 7 measurement	58
Minimum 10 measurement	51
Minimum 13 measurement	31

6.1.2. Izmir Dataset

The Izmir dataset consists of 383 measurements taken from different parts of the city. The region and the fingerprints are shown in Figure 6.4. The TA and Rx histogram of the measurements are shown in Figure 6.5 and Figure 6.6, respectively. To show the effect of number of measurements for fingerprinting, the dataset is divided into several regions

according to their serving BTS ids. The table 6.2 shows the number of the measurements for the scenarios of Izmir dataset.

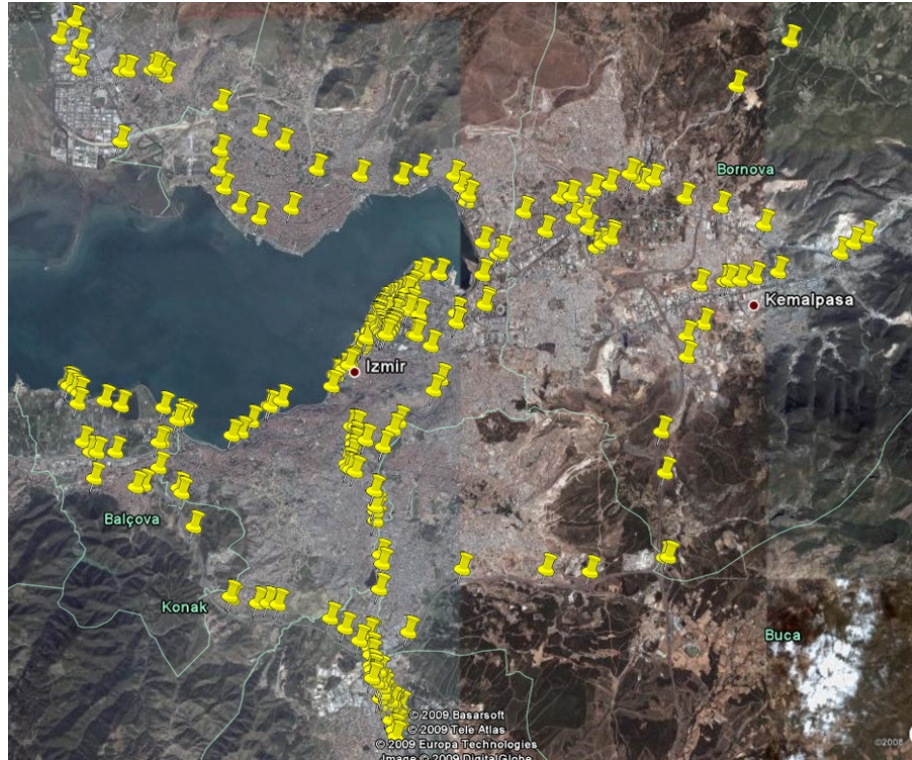


Figure 6.4. Izmir measurements

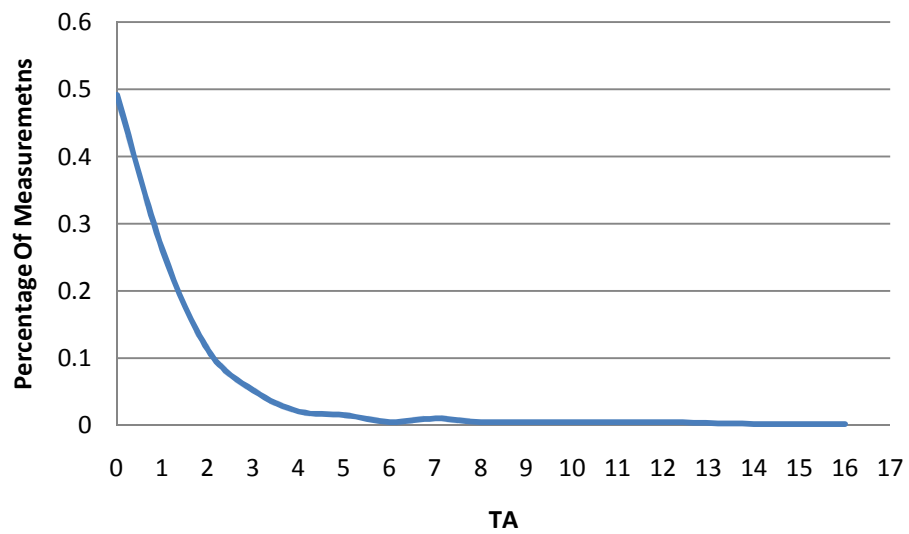


Figure 6.5. TA histogram of Izmir measurements

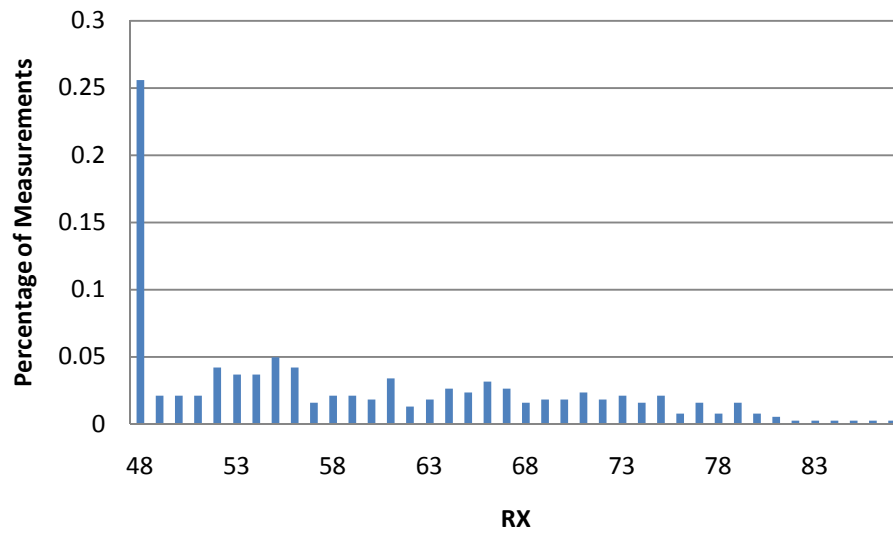


Figure 6.6. Rx histogram of Izmir measurements

Table 6.2. Number of measurements for same serving BTS in Izmir

	Number of Measurements
Minimum 1 measurement	340
Minimum 2 measurement	278
Minimum 3 measurement	208
Minimum 4 measurement	193
Minimum 5 measurement	173
Minimum 6 measurement	158
Minimum 7 measurement	146
Minimum 9 measurement	139
Minimum 10 measurement	121
Minimum 12 measurement	101
Minimum 19 measurement	89
Minimum 31 measurement	70
Minimum 39 measurement	39

6.2. Threshold and Coefficient Calculation

Euclidean Distance (ED), Manhattan Distance (MD), and DCM for Multi-System Location (MSL) methods are based on a threshold value. The threshold value is used to calculate the punishment value of a cell not accessible from either the MS or the fingerprint in the database. There is no threshold value for DCM for GSM location method. Our method uses a coefficient value for database correlation. ED, MD and MSL use a threshold value to calculate the punishment part of respective algorithms. The threshold values are chosen to minimize the error. The values between 0 and 200 are tested for the threshold and the value that gives the minimum error is chosen for different scenarios.

For our method, the coefficient value is calculated to minimize the error. Values between 0 and 10000 are tested to find the one that minimizes error value for different scenarios.

Table 6.3. Threshold and Coefficient Values for two Measurement Sets

	Istanbul	Izmir
Euclidean Distance	84	12
Manhattan Distance	72	124
DCM Multi-System Location	84	12
LEF	1443	1416

6.3. Performance Comparison

The performances of the ED, MD, and MSL are compared against LEF. The results are obtained by averaging ten runs. The results show that LEF method decreases the average error for Istanbul measurements by approximately 10% and for Izmir measurements 0.5%.

In Figures 6.7 and 6.9, the effect of number of measurements under the same BTS is shown. Istanbul and Izmir measurements are divided into groups according to their number of measurements under the same serving BTS. Then, the experiments are repeated for

those measurements ten times and the averages are taken. The experiments show that the average error decreases when the number of measurements are increased. The results show that LEF provides the minimum error in distance for both Istanbul and Izmir datasets.

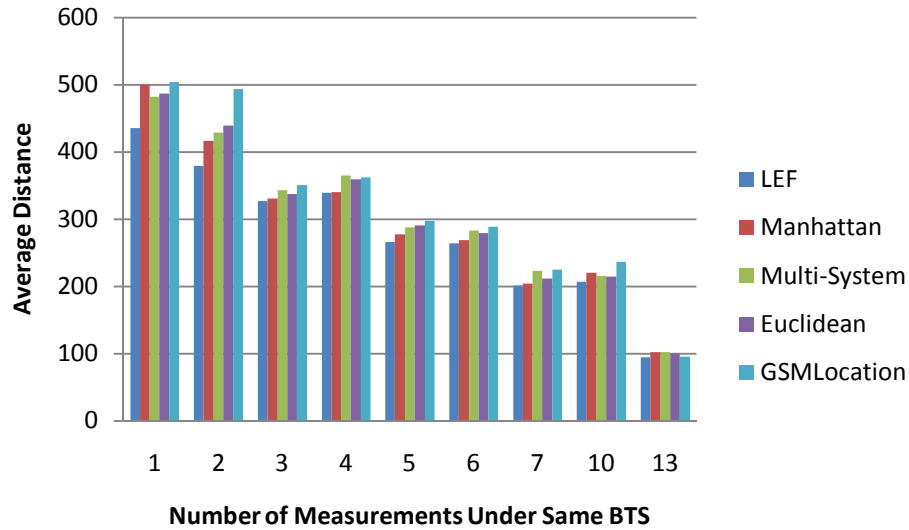


Figure 6.7. Average error in distance for Istanbul

In Figures 6.8 and 6.10, the CDF comparisons of the algorithms for Istanbul and Izmir are shown. For Istanbul dataset, there is a calculation which gives an error around 2500m. For that maximum error point, LEF gives the best result among others. When we focus on the CDF, LEF obtains 67% and 95% of the measurements with the minimum error. For Izmir dataset, the error is around 8000m. For that single point, only GSMLocation gives better result than LEF. On the other hand, as can be seen in the CDF graph, although GSMLocation gives better result for that specific point, LEF reaches 67% and 95% better than the other algorithms. Additionally, GSMLocation has four calculations that have error above 2000m while the other algorithms maintain an error around 8000m.

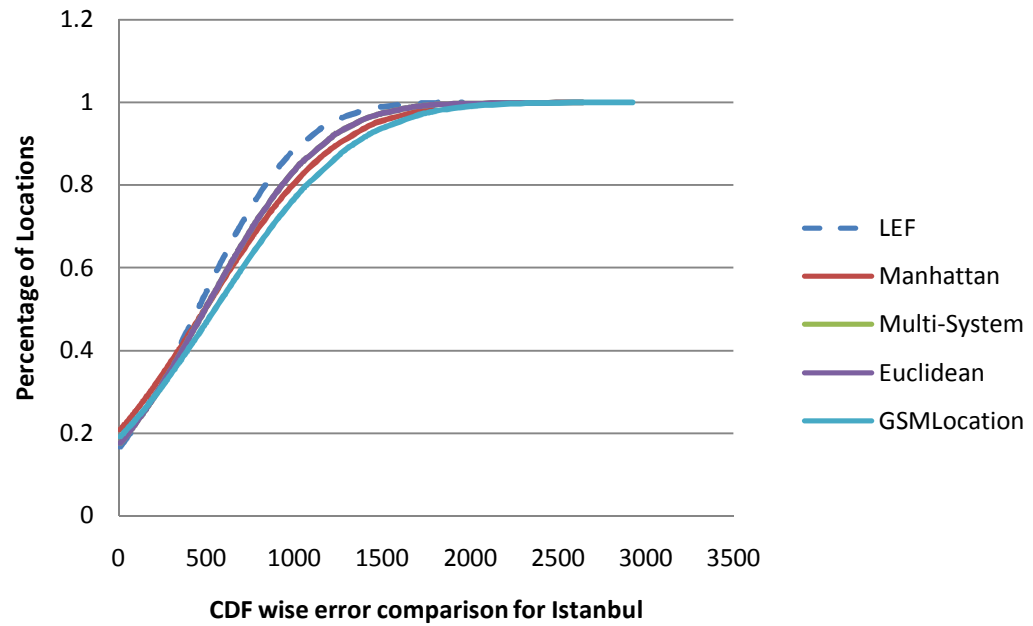


Figure 6.8 CDF comparison for Istanbul Measurements

Table 6.4 Mean comparison for Istanbul

	LEF	Manhattan	Multi-System	Euclidean	GSMLocation
Average	435.0209	500.3306	481.6278	486.3379	504.2638

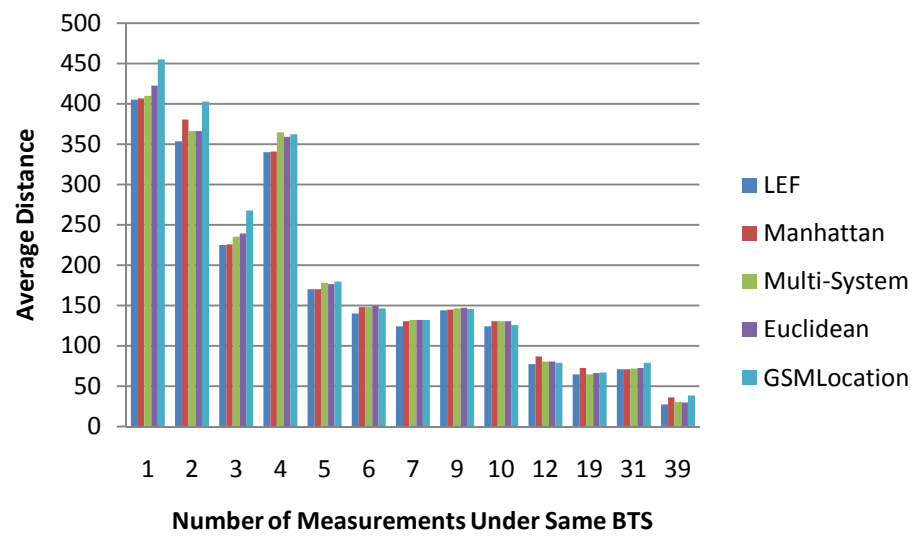


Figure 6.9. Average error in distance for Izmir

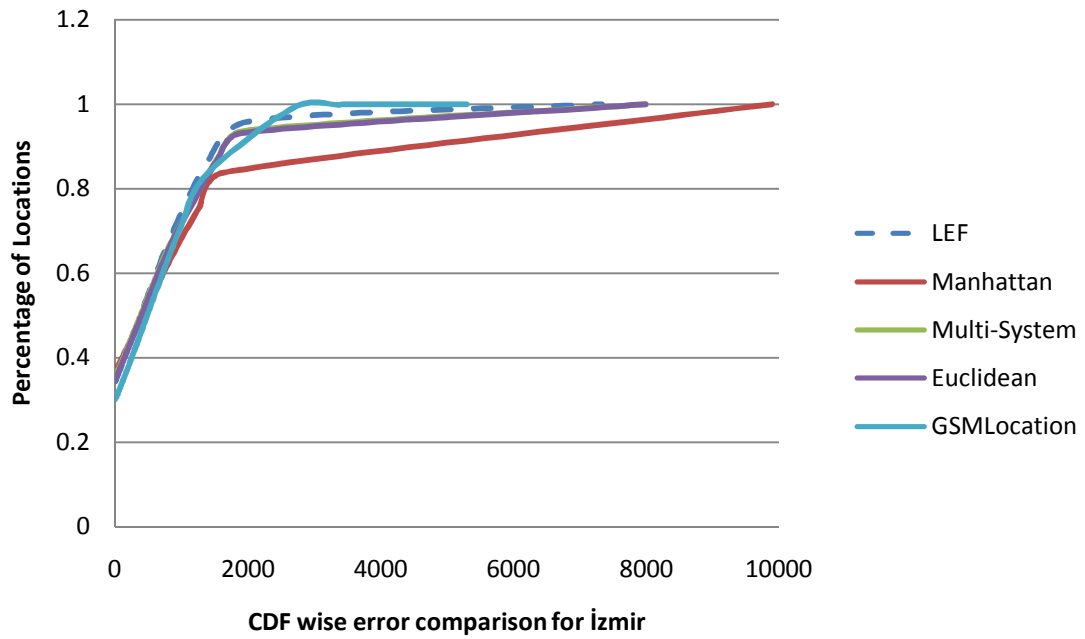


Figure 6.10 CDF comparison for Izmir

Table 6.5 Mean comparison for Izmir

	LEF	Manhattan	Multi-System	Euclidean	GSMLocation
Average	404.8652	406.5909	409.7465	422.0976	454.7081

6.4. Number of Nearest Neighbor Calculation

The error in distance is calculated for k values from one to five. When there is only one neighbor, the distance is calculated by Vincenty's formula [35]. For more than one neighbors, WkNN algorithm is used. The reason behind using WkNN is its compliance to the location estimation by database correlation method. With database correlation method, the correlations between the MS and the fingerprints are calculated. These correlations correspond to the weights in WkNN. For more than one neighbors, a new point is calculated by using the correlations as weight and the fingerprints coordinates. Let x be the x -coordinate of the MS and y be the y -coordinate of the MS, and F_1, F_2, F_3 be the fingerprints that have the minimum correlations. In addition, x_1, x_2, x_3 are the x -coordinates, y_1, y_2, y_3 are y -coordinates and c_1, c_2, c_3 are the correlations of the F_1, F_2, F_3 respectively.

$$c_t = c_1 + c_2 + c_3 \quad (6.1)$$

$$x_{calc} = x_1 \times \frac{\frac{1}{c_1}}{\frac{1}{c_t}} + x_2 \times \frac{\frac{1}{c_2}}{\frac{1}{c_t}} + x_3 \times \frac{\frac{1}{c_3}}{\frac{1}{c_t}} \quad (6.2)$$

$$y_{calc} = y_1 \times \frac{\frac{1}{c_1}}{\frac{1}{c_t}} + y_2 \times \frac{\frac{1}{c_2}}{\frac{1}{c_t}} + y_3 \times \frac{\frac{1}{c_3}}{\frac{1}{c_t}}. \quad (6.3)$$

By generalizing equation 6.1, 6.2, and 6.3,

$$c_t = \sum_{n=1}^k c_n \quad (6.4)$$

$$x_{calc} = \sum_{n=1}^k x_n \times \frac{\frac{1}{c_n}}{\frac{1}{c_t}} \quad (6.5)$$

$$y_{calc} = \sum_{n=1}^k y_n \times \frac{\frac{1}{c_n}}{\frac{1}{c_t}} \quad (6.6)$$

$$CP(x_{calc}, y_{calc}) \quad (6.7)$$

$$dist = Vincenity(x, x_{calc}, y, y_{calc}) \quad (6.8)$$

where CP is the calculated point and Vincenity is the distance calculation formula. For more than one neighbor, the distance is calculated by using the equations 6.4, 6.5, 6.6, 6.7, and 6.8.

For the determination of the number of neighbors that give the best one to five is tried for all experiments. They give their best results for different k values. For Istanbul measurements, the methods give their best results with only one neighbor. For Izmir measurements, the methods give their best results for three neighbors.

Table 6.6. Optimum k value for Istanbul and Izmir

	Number of Measurement	k value for WkNN
Istanbul	375	1
Izmir	340	3

For the Istanbul dataset, the reason of reaching the minimum error with one neighbor is the sampling rate. As observed in Figure 6.1, the measurements are taken on the road and the distances between the measurements are not small and the measurements are directional. This means that if the nearest neighbors cannot form a triangle in which the MS stands, applying the WkNN may increase the error instead of decreasing.

For the Izmir dataset, the situation is different from Istanbul. The bay area and the south area of the measurements have smaller sampling rate. The fingerprints in these areas are close to each other. So, applying WkNN decreases error in distance compared to the one neighbor and two neighbor cases. Although four and five neighbors are tested, the best result are achieved with three neighbors. This can be explained with the correlation calculation. LEF and the other algorithms calculate the correlations according to the Rx values which loss their magnitude depending to the environment. So, the correlation calculations may give good results for the nearest neighbors up to a number. When that limit is passed, the correlations begin to lose correctness due to effects described in

Chapter 2. When the correlations lose correctness, the error in WkNN begins to increase instead of decreasing. So, a k value should be determined for the environment. This can be done by tests and the characteristics of the environment define the k value which is three for Izmir dataset.

7. CONCLUSIONS

Location estimation offers a wide range of new services to the users and operators. With the information of location, new services and the current services will become more powerful. The services such as advertising, dating, games, media, billing and emergency services will be affected in a good way with the developments in the LBS techniques.

Accuracy of the location based services is a crucial issue. Fingerprinting method may give accurate results according to its implementation and the fingerprint database. Additionally, Fingerprinting method can be applied to any cellular network. In this work, we have tried to propose a new fingerprint matching algorithm to increase the accuracy of the fingerprinting method.

We have presented our LEF fingerprint matching algorithm and all the other phases used during the location estimation process. We have shown that the proposed method enhances the current algorithms in average error. The key feature of the LEF is that it can be used in any Fingerprinting mechanism with minimal cost on the legacy infrastructure. It is relatively lightweight and necessitates only a small amount of preliminary work for the coefficient calculation.

REFERENCES

1. Rappaport, T. S., J. H. Reed and B. D. Woerner, "Position Location Using Wireless Communications on Highways of Future," *IEEE Communications Magazine*, Vol. 34, Issue 10, pp. 33-41, 1996.
2. Adusei, I.K., K. Kyamakyaa 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," *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. Knuuttilla, "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-sight environments," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Atlanta, USA, pp. 274-278, 2004.
11. Hata, M., "Empirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Trans. Veh. Technol.*, Vol. 29, no. 3, pp. 317-325, 1980.
12. 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.
13. Saunders S. R., Zavala A. A., "Antennas and Propagation for Wireless Communication Systems", 2nd Edition , John Wiley & Sons, 2007.
14. Parsons, J. D., "The Mobile Radio Propagation Channel", 2nd Edition, John Wiley & Sons, 2000.
15. Tonteri, T., "A Statistical Modeling Approach to Location Estimation", *M.S. Thesis*, University of Helsinki, 2001.
16. Kozono S., Watanabe K., "Influence of environmental buildings on UHF land mobile radio propagation", *IEEE Transactions on Commununications*, 1977.

17. WG 2, "Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands", Technical Report COST 231 TD (90) 119 Rev. 2, 1990.
18. 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.
19. Türkylmaz O., *Environment Aware Location Estimation in Cellular Networks*, M.S. Thesis, Boğaziçi University, 2007.
20. Sayed, A. H. A. Tarighat and Khajenouri, "Network Based Wireless Location: Challenges Faced in Developing Techniques of GSM Phones with the Simulations", *IEEE Signal Processing Magazine*, Vol. 22, July 2005.
21. Schiller J. H., Voisard A., *Location-based services*, Morgan Kaufmann, 2004.
22. Kolodziej K. W., Hjelm J., *Local positioning systems*, CRC Press, 2006.
23. LaMance, J., J. Jarvinen and J. DeSalas, "Assisted GPS: A Low-Infrastructure Approach," *GPS World*, 2002.
24. Laitinen, H., S. Ahonen, S. Kyriazakos, J. Lähteenmäki, R. Menolascino and S. Parkkila, *Cellular Network Optimisation Based on Mobile Location*, Technical Report, 2001.
25. W. G. Figel, N. H. Shepherd, and W. F. Trammel, "Vehicle Location by a Signal Attenuation Method," *IEEE Transactions on Vehicular Technology*, 18(3):105-109, November 1969.
26. Laitinen H., Lahteenmaki J., Nordstrom T., "Database Correlation Method for GSM Location," *IEEE Vehicular Technology Conference*, Vol. 4, pp. 2504-2508, 2001.

27. Lakmali B.D.S., Dileeka. D., “Database Correlation for GSM Location in Outdoor & Indoor Environments,” *Information and Automation for Sustainability, ICIAFS*, pp. 42-47, 2008.
28. Lakmali B.D.S., Wijesinghe W.H.M.P., De Silva K.U.M., Liyanagama K.G., Dias S.A.D., “Design, Implementation & Testing of Positioning Techniques in Mobile Networks,” *Information and Automation for Sustainability, ICIAFS*, pp. 94-99, 2007.
29. Kemppi P., “Database Correlation Method For Multi-System Location,” *M.S. Thesis*, Helsinki University of Technology, 2005.
30. Kamol K., "Design of Indoor Positioning Systems based on Location Fingerprinting Technique", Ph.D. Dissertation, The School of Information Science, University of Pittsburgh, 2005.
31. Djuknic G.M., Richton R.E., "Geo-location and assisted GPS," *IEEE Computer*, Vol. 34, pp. 123–125, Feb. 2001.
32. 3rd Generation Partnership Project Technical Specification Group, “3GPP TS 45.010”, Technical Specification, Radio subsystem synchronization, V8.3.0, February 2009.
33. Li B., Wang Y., Lee H. K., Dempster A., and Rizos C., “A New Method of Yielding a Database of Location Fingerprints in WLAN,” *Communications, IEE Proceedings*, Vol. 152, pp. 580-586, 2005.
34. Pandya D., Jain R., Lupu E., “Indoor location estimation using multiple wireless technologies,” *Personal, Indoor and Mobile Radio Communications, PIMRC 2003. 14th IEEE Proceedings*, Vol. 3, pp. 2208-2212, 2003.
35. Vincenty T., “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations,” *National Geodesic Survey Review XXII*, pp. 176, 1975.

36. Yamamoto R., Matsutani H., Matsuki H., Oono T., and Ohtsuka H., "Position Location Technologies using signal strengths in cellular system," *Vehicular Technology Conference, IEEE VTS 53rd.*, Vol. 4, pp. 2570-2574, 2001.
37. Takenga C., Kyamakya K., "Location Fingerprinting in GSM network and Impact of Data Pre-processing," *WMC06*, 2006.
38. Khalaf-Allah M., Kyamakya K., "Database Correlation using Bayes Filter for Mobile Terminal Localization in GSM Suburban Environments," *Vehicular Technology Conference, VTC 2006 IEEE 63rd*, Vol. 2, pp. 798-802, 2006.
39. Takenga C., Xi C., Kyamakya K., "Fusion of Neural Network positioning and Database Correlation in localizing a Mobile Terminal," *IEEE International Conference on Wireless Networks (ICWN'06)*, pp. 26-29, 2006.
40. Klukas R. W., "A Super resolution Based Cellular Positioning System Using GPS Time Synchronization", *Ph.D. Dissertation*, Department of Geomatics Engineering ,The University of Calgary, Alberta, 1997.
41. Siegman A. E., "The Antenna Properties of Optical Heterodyne Receivers," *Applied Optics*, Vol. 5, pp. 1588-1594, 1966.
42. Dru M. A., Saada S., "Location-based mobile services: the essentials", *Alcatel Telecommunications Review*, 2001.
43. Takenga C., Xi C., Kyamakya K., "A Hybrid Neural Network-Data Base Correlation Positioning in GSM Network," *Communication systems, ICCS 2006, 10th IEEE Singapore International Conference on*, pp. 1-5, 2006.
44. Türkyilmaz O., Alagoz F., Gur G., Tugcu T., "Environment-aware location estimation in cellular networks", *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, 2008.

45. Wijesinghe P., Dias D., “Novel Approach for RSS Calibration in DCM -based Mobile Positioning using Propagation Models,” *Information and Automation for Sustainability, ICIAFS 2008, 4th International Conference on*, pp. 29-34, 2008.
46. Salsic Z., Chan E., “Mobile station positioning using gsm cellular phone and artificial neural networks,” *Wireless Personal Communications*, Vol. 14, pp. 235–254, 2000.
47. Bajada J., “Mobile positioning for location dependent services in gsm networks,” *MobilePositioning.pdf*. [Online]. Available: <http://www.cs.um.edu.mt/~csaw/CSAW03/Proceedings>
48. Triki M., Slock D., Rigal V., Francois P., “Mobile terminal positioning via power delay profile fingerprinting: Reproducible validation simulations,” *Vehicular Technology Conference, VTC-2006 Fall, 2006 IEEE 64th*, pp. 1–5, 2006.
49. Zimmermann D., Baumann J., Layh A., Landstorfer F., Hoppe R., Wolfle G., “Database correlation for positioning of mobile terminals in cellular networks using wave propagation models,” *Vehicular Technology Conference, VTC2004-Fall, 2004 IEEE 60th*, Vol. 7, pp. 4682–4686, 2004.
50. Wijesinghe P., Dias D., “Novel Approach for RSS Calibration in DCM -based Mobile Positioning using Propagation Models,” *Information and Automation for Sustainability, ICIAFS 2008, 4th International Conference on*, pp. 29-34, 2008.