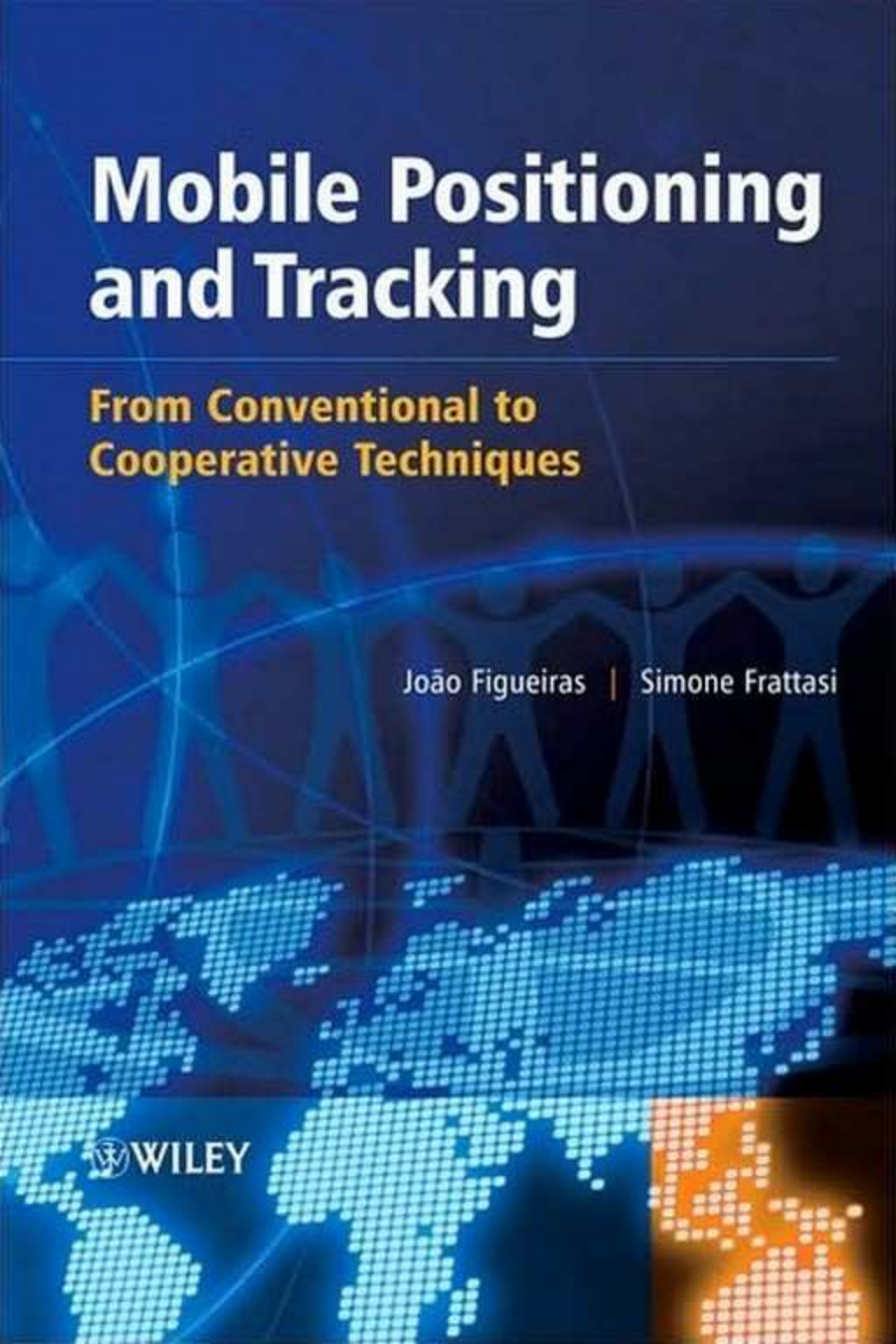
Mobile Positioning and Tracking

From Conventional to Cooperative Techniques

João Figueiras | Simone Frattasi



WILEY

Mobile Positioning and Tracking

Mobile Positioning and Tracking

From Conventional to Cooperative Techniques

João Figueiras
Aalborg University, Denmark

Simone Frattasi
Aalborg University, Denmark

WILEY
A John Wiley and Sons, Ltd, Publication

This edition first published 2010
© 2010 John Wiley & Sons Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Figueiras, João.
Mobile positioning and tracking : from conventional to cooperative techniques / João Figueiras, Simone Frattasi.
p. cm.
Includes bibliographical reference and index.
ISBN 978-0-470-69451-0 (cloth : alk. paper)
1. Location-based services. 2. Mobile geographic information systems. 3. Wireless communication systems. 4. Electronics in navigation. I. Frattasi, Simone. II. Title.
TK5105.65.F54 2010
621.384–dc22
2010008337

A catalogue record for this book is available from the British Library.

ISBN 978-0-470-69451-0 (H/B)

Set in 11/13pt Times by Sunrise Setting Ltd, Torquay, UK.
Printed and Bound in Singapore by Markono Print Media Pte Ltd.

. . . to my family

João Figueiras

. . . to my love, Sara

Simone Frattasi

Contents

About the Authors

João Figueiras received his Ph.D. in Wireless Communications from Aalborg University, Aalborg, Denmark, in 2008 and his *Licenciatura* in Electrical Engineering and Computer Science from the Instituto Superior Técnico, Technical University of Lisbon (IST-UTL), Portugal, in 2004. From 2004 to 2008, he was employed by Aalborg University to work on the Danish-funded projects WANDA (Wireless Access Networks, Devices and Applications) and COMET (Cooperative Mobile Positioning), and the European-funded projects MAGNET (My Personal Adaptive Global Net) Beyond and WHERE (Wireless Enhanced Mobile Radio Estimators). His work was in the area of positioning and tracking solutions for wireless networks, and his results have been used by several of the project partners, such as Blip Systems and Deutschen Zentrums für Luft- und Raumfahrt (DLR). His research has been published in several papers and presented at conference workshops (at IEEE PIMRC'07). During 2007, he was a research visitor at the Electrical Engineering Department of UCLA, Los Angeles, USA, and an intern software engineer in the Mobile Team at Google Inc., Mountain View, USA. Since 2009 he has been employed by ZTE Portugal as the after-sales manager for the mobile market in Portugal. From 2003 and in parallel with his professional activities, he has been privately and continuously involved in several innovation projects in the area of cooperative communications and Internet services. Since 2006 he has been an active volunteer of IEEE, where he currently holds the position of IEEE Region 8 (EMEA) GOLD (Graduates of Last Decade) committee chair; his main duties are to increase the involvement of members in IEEE activities, coordinate communication among IEEE young members and leverage the benefits of IEEE in order to meet the needs of young members' careers. He has been the main organizer of conferences (AISPC'07 and AISPC'08) and worldwide IEEE meetings (1st World GOLD Summit 2008).

Simone Frattasi received his Ph.D. in Wireless Communications from Aalborg University, Aalborg, Denmark, in 2007, and his M.Sc. degree *cum laude* and his B.Sc. degree in Telecommunications Engineering from Tor Vergata University, Rome, Italy, in 2002 and 2001, respectively. Since 2009 he has worked as a patent attorney in the intellectual property consultancy company Plougmann & Vingtoft (P&V), Århus, Denmark. Before joining P&V, he was employed from 2002 to 2005 as a research assistant at Aalborg University, where he worked on two European projects (STRIKE and VeRT) and one industrial project (JADE) in collaboration with the Global

Standards & Research Team, Samsung Electronics Co. Ltd., Korea. From 2005 to 2007 he was an assistant professor and, besides still contributing to the JADE project, he led a Danish-funded project named Cooperative Mobile Positioning (COMET). From 2007 to 2008 he was the project leader of an industrial project (LA-TDD) in collaboration with Nokia Siemens Networks, Aalborg, Denmark. Dr. Frattasi is the author of numerous scientific and technical publications. He has served as a technical programme committee chairman for several conferences and as a guest editor for several magazines and journals. Moreover, he was the main instructor for a half-day tutorial on wireless location at IEEE PIMRC'07. Simone is a co-founder of the International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART) and the International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL). Finally, he had the honor to be appointed as a chairman of IEEE GOLD AG Denmark in 2008.

Preface

Localization is a research topic that is receiving increasing attention from both academia and industry. Previously considered as vital information for vehicle tracking and military strategy, location information has now been introduced into wireless communication networks. In contrast to dedicated solutions, such as the Global Positioning System (GPS), that were designed to simply provide positioning information, the new solutions for wireless networks are able to supply the combined benefit of both communication and positioning. As a consequence, the network operator, as well as the service provider and the end user, can profit from such position-enabled communication capabilities. Indeed, while the network operator is able to manage the resources of its network more efficiently, the service provider is able to offer location-based services (LBSs) to the end user, who can fully enjoy such personalized location-dependent services. In particular, it can be found from the literature that location information is being used as a basic requirement for the deployment of new protocols (e.g., routing and clustering), new technologies (e.g., cooperative *systems*) and new applications (e.g., navigation and location-aware advertising). From the point of view of the industry, the use of location information has been stimulated mainly by applications such as navigation, location-dependent searching and social networking. Since wireless communication networks are nowadays present anywhere and anytime, every location-dependent networking enhancement, service or application can be spread rapidly and used globally.

The above-mentioned trends are a major stimulator for the development of novel solutions for obtaining positioning information in wireless networks. Chapter 1 outlines the motivation behind these solutions and presents potential categories and applications of location-based services (both conventional and network-related). Chapter 2 introduces the basic notions of wireless communications needed to ease the reader through the remaining chapters of the book. Chapter 3 presents the fundamentals of positioning, from types of measurements and estimation techniques to error sources and typical metrics for location accuracy. Owing to the intrinsic nature of wireless systems, obtaining a position fix for a user requires the use of robust algorithms for associating data retrieved from several sources at different moments in time. Chapter 4 describes these various types of data association algorithms, showing the advantages and disadvantages of each of them. Chapter 5 deals with the fundamentals of tracking; in particular, several mobility models (including group-based and socially-based models) to be used in Chapter 8 will be introduced.

Chapter 6 considers some advanced techniques (from the realm of signal processing) used to mitigate the errors mentioned in Chapter 5, thus trying to enhance the accuracy of the overall location estimation process. Chapter 7 presents the state of the art of satellite-based and terrestrial based positioning systems, spanning the range from outdoor to indoor environments, from wide-area networks to short-range networks, and from orthogonal frequency division multiplexing (OFDM) to ultra-wideband (UWB) technologies. Recently, we have observed that replicating cooperative human behavior in wireless communications has resulted in a number of emerging research fields. In particular, its application in wireless location has flown in a new breed of techniques that may revolutionize the entire field. Hence, in Chapter 8, we take a tour through the state of the art of what we call "cooperative augmentation systems", i.e., mobile positioning systems that exploit the cooperation of users, terminals and networks to boost their location estimation accuracy.

This book includes an accompanying website. Please visit `www.wiley.com/go/figueiras_mobile` for more information.

Acknowledgements

The authors would like to thank the direct contributors of the book, namely Nicola Marchetti, Ismail Guvenc, Andreas Waadt, Guido H. Bruck and Peter Jung, and the indirect contributors, who, in one way or another, have been involved in several of the activities that provided the know-how for writing the book: Hans-Peter Schwefel, Ramjee Prasad, Ali Sayed, Cassio Lopes, Adel Youssef, Rasmus Olsen, Gian Paolo Perrucci, Istvan Kovacs, Lars Tørholm Christensen, Marco Monti, Francescantonio Della Rosa and Gianluca Simone. Finally, the authors would like to thank the Wiley team, who offered them their unceasing help in order to make this book a reality: Tiina Ruonamaa, Sarah Tilley, Anna Smart and Alistair Smith.

List of Abbreviations

1L-DF	one-level data fusion
2D	two-dimensional
2L-DF	two-level data fusion
3D	three-dimensional
3G	third generation
3GPP	Third Generation Partnership Project
4G	fourth generation
A-GNSS	Assisted Global Navigation Satellite System
A-GPS	Assisted Global Positioning System
ADC	analog-to-digital converter
AOA	angle of arrival
AP	access point
AWGN	additive white Gaussian noise
B2C	Business to Consumer
BF	beamforming
BPSK	binary phase shift keying
BS	base station
BTS	base transceiver station
CA	collision avoidance
CAS	cooperative augmentation system
CD	collision detection

CDF	cumulative distribution function
CDMA	code division multiple access
CEP	circular error probability
CH	cluster head
CID	cell ID
CIR	carrier-to-interference ratio
CLI	caller location information
CLS	constrained least squares
CM	cluster member
COFDM	coded orthogonal frequency division multiplexing
COMET	Cooperative Mobile Positioning System
coop-EKF	cooperative extended Kalman filter
coop-WNLLS	cooperative weighted nonlinear least squares
CR	cognitive radio
CRC	cyclic redundancy check
CRLB	Cramér–Rao lower bound
CRMSE	cooperative root mean square error
CSMA	Carrier sense multiple-access
CTM	current transformation matrix
CTS	clear to send
CW	continuous-wave
DAC	digital-to-analog converter
DGPS	Differential Global Positioning System
DOP	dilution of precision
DR	dead reckoning
DSSS	direct-sequence spread spectrum
E911	Enhanced 9-1-1

eCall	Emergency Call
EGNOS	European Geostationary Navigation Overlay Service
EKF	extended Kalman filter
EM	expectation maximization
EPS	evolved packet system
ERP	equivalent radiated power
EUWB	European Ultra-Wideband
FCC	Federal Communications Commission
FDMA	frequency division multiple access
FEC	forward-error-correction
FFT	fast Fourier transform
FHSS	frequency-hopping spread spectrum
FIM	Fisher information matrix
FRP	fixed reference point
G-CRLB	generalized Cramer–Rao lower bound
GAGAN	GPS-Aided Geo-Augmented Navigation system
GDOP	geometric dilution of precision
GLONASS	Globalnaya Navigationnaya Sputnikovaya Sistema
GPB	generalized pseudo-Bayesian
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HazMat	hazardous material
HLOP	Hybrid Lines of Position
HTAP	Hybrid TOA/AOA Positioning
HTDOA	Hybrid TDOA/AOA Positioning
IAD	identify and discard

IEEE	Institute of Electrical and Electronics Engineers
IFFT	inverse fast Fourier transform
IP	Internet Protocol
IPO	interior-point optimization
IS	idle slot
IT	information technology
KF	Kalman filter
LAC	local area code
LAN	local area network
LBS	location-based service
LCS	location services
LD	laser diode
LED	light emitting diode
LLS	linear least-squares
LMS	least-median-of-squares
LMU	location measurement unit
LOS	line-of-sight
LP	local-positioning
LS	least-squares
LT	location and tracking
LTE	long-term evolution
LTS	least-trimmed-squares
MAC	medium access control layer
MCC	mobile country code
MIMO	multiple-input–multiple-output
MISO	multiple-input–single-output
ML	maximum-likelihood

MLE	maximum-likelihood estimator
MNC	mobile network code
MS	mobile station
MSAS	Multifunctional Satellite Augmentation System
MSC	mobile switching center
MSE	mean squared error
NAV	network allocation vector
NAVSTAR	Navigational Satellite Timing and Ranging
NLLS	nonlinear least-squares
NLOS	non-line-of-sight
NNSS	Navy Navigation Satellite System
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OSI	Open System Interconnection
OTDOA-IPDL	Observed Time Difference of Arrival–Idle Period Downlink
P2P	peer-to-peer
PC	power control
PCR	predictive-channel-reservation
PDF	probability density function
PF	particle filter
PHY	physical layer
PLMN	public land mobile network
PRN	pseudo-random number
PSAP	public safety answering point
PSTN	public switched telephone network
QAM	quadrature amplitude modulation
QoS	quality-of-service

QP	quadratic programming
QPSK	quadrature phase shift keying
QZSS	Quasi-Zenith Satellite System
RBF	recursive Bayesian filtering
RF	radio frequency
RFID	radio frequency identification
RLS	recursive least-squares
RMS	root mean square
RMSE	root mean square error
RRM	radio resource management
RSS	received signal strength
RSSI	received signal strength indicator
RT	residual test
RTD	relative time difference
RTS	request to send
RX	receiver
SBAS	Satellite-Based Augmentation System
SD	spatial diversity
SDMA	space division multiple access
SIMO	single-input–multiple-output
SINR	signal-to-interference-plus-noise ratio
SISO	single-input–single-output
SNR	signal-to-noise ratio
STC	space–time coding
TA	timing advance
TDMA	time division multiple access
TDOA	time difference of arrival

THSS	time-hopping spread spectrum
TOA	time of arrival
TX	transmitter
U-TDOA	uplink–time difference of arrival
UKF	unscented Kalman filter
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
UWB	ultra-wideband
WAAS	Wide Area Augmentation System
WGS84	World Geodetic System 1984
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	wireless local area network
WLS	weighted least squares
WNLLS	weighted nonlinear least squares
WPAN	wireless personal area network

Notations

Symbol	Description
x, y, a	scalar
X, Q, R	vector or matrix
α, μ, θ	scalar, vector or matrix
X_{rel} or X^{rel}	X with a descriptive label "rel"
X^{T}	transpose of X
X^{-1}	inverse of X
$\hat{X}$	estimator of X
$\tilde{X}$	$\hat{X} - X$
$\bar{X}$	mean of X
$\lvert X \rvert$	determinant of X
$X^{(i)}$	X with respect to element i
$X^{(i)(j)}$	X with respect to a relation between i and j, both from the same group
$X^{(i)[j]}$	X with respect to a relation between i and j from a different group
X_i	equivalent to $X^{(i)}$
$X_{i,j}$	equivalent to $X^{(i)(j)}$
X_k or $X(k)$	X at discrete time k
$X_{k\lvert l}$	X at discrete time k compared with time l
$X_{l:k}$	vector or matrix containing all occurrences of X from discrete time l until k
Δx	difference between two values of x
$\binom{n}{i}$	set of i-combinations of a set n
$\mathrm{P}(x)$	probability of x
$\mathrm{f}(x)$	PDF of x
$\mathrm{Norm}(\mu, \sigma)$	normal distribution with mean μ and standard deviation σ
$\mathrm{Unif}(a, b)$	uniform distribution between a and b
$\mathrm{Exp}(\bar{\gamma})$	exponential distribution with parameter $\bar{\gamma}$

Symbol	Description
Tri(a, b)	triangular distribution with lower limit a, upper limit b and center $(a+b)/2$
log(•)	logarithm with base 10
ln(•)	logarithm base e (neper number)
f	frequency
λ	wavelength
c	speed of light
p	power
α	constant term in the path loss or received power equation
β	path loss exponent
t	time
t	time-difference
d	distance
d	difference between distances
E[•]	expected value
η	noise component
μ	mean value
σ	standard deviation
θ	angle
γ	weight value
i or j	iteration constant
x or X	position coordinate
z or Z	measured value
$\mathcal{H}(\bullet)$	function relating position coordinates and measured values
H	Jacobian of $\mathcal{H}(\bullet)$
$\mathcal{A}(\bullet)$, $\mathcal{F}(\bullet)$	function relating position coordinates at different points in time
A	Jacobian of $\mathcal{A}(\bullet)$
∇	gradient operator
$\mathcal{J}(\bullet)$	cost function in an optimization procedure

This list normalizes the notation throughout the book, however, at selected and well identified parts a different notation may be used for convenience purposes.

1

Introduction

Over the past few decades, wireless communications have become essential in everyone's daily life. The world has become mobile, and continuous access to information has become a requirement. Owing to this necessity for fresh, real-time, first-hand information, devices such as mobile phones, computers, pagers, data cards, sensors and data chips have been entering our lives as typical technology "buddies". For this reason, wireless services have gained popularity, and location information has become useful information in the wireless world. The continuous demand for information has created a huge potential for business opportunities and it has promoted innovation towards the development of new services. As a consequence, the infrastructure that enables communication among wireless devices has been rapidly growing towards higher coverage, higher flexibility and higher interoperability. This reality is so visible that it is nowadays unthinkable to envision our lives without these technologies. Furthermore, with the rapid deployment of wireless communication networks, positioning information has become of great interest. Because of the inherent mobility behavior that characterizes wireless communication users, location information has also become crucial in several circumstances, such as in rescues, emergencies and navigation. This position dependency has boosted research, development and business around the topic of positioning mechanisms for wireless communication technologies. The result is a wide variety of integrated and built-in solutions that can combine and interoperate with communication and position information. Thus, this book covers the topic of positioning mechanisms for wireless communication technologies. It explains in detail the services, wireless communication protocols, positioning and tracking algorithms, error mitigation techniques, implementations used in wireless communication systems, and the most recent techniques of cooperative positioning.

Positioning or location can be understood as the unambiguous placement of a certain individual or object with respect to a known reference point. This reference point is often assumed to be the center of the Earth coordinate system. In practice,

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

the reference point can be any point on the Earth[1] that is known to the system and which all the coordinates can relate to. Although the position itself is obviously a very important source of information, this position must be related to a specific time to be even more useful. In particular, when we consider tracking systems, time information is a key necessity, not only for knowing the position of a certain device at a specific time, but also for inferring higher-order derivatives of the position, i.e., speed and acceleration. Thus, "tracking" is a method for estimating, as a function of time, the current position of a specific target. "Navigation" is a tracking solution that aims primarily at using position information in order to help users to move towards a desired destination.

Although enabling position estimation with communication technologies is currently a hot topic, the necessity for determining the position of individuals, groups, animals, vehicles or any type of object is an ancient necessity that can be seen as a basic need. This necessity has been present in human life for many centuries and it is so important that humans and animals have in-built biological mechanisms that permit individuals to localize and orient themselves in many situations. When we consider the actual methods for obtaining position information, the history is long and the systems are numerous. During the early years, a few millennia ago, orientation and positioning were already possible using devices that resemble present-day magnetic compasses, using maps of sea currents and winds, or using celestial navigation techniques. For centuries, celestial navigation, by means of observations of the positions of stars and the Sun, was the most important technique for estimating position information. By knowing, for instance, the position of the Sun or well-known star constellations, people were able to know their orientation and navigate on the Earth. This technique is so important that it is still used today in order to provide a rough sense of one's orientation when no other tool is available. Although the mechanisms for orientation based on star position readings and compass readings have been used for at least two millennia, these mechanisms were widely used and improved during the period of sea exploration in the 15th and 16th centuries. This was an important period in the history of positioning systems. Several objects, such as the cross-staff and astrolabe (and, later, the quadrant and the sextant, invented in the 17th and 18th centuries), permitted sea explorers to read the position of the stars and subsequently calculate their own position, often supported by the incomplete maps that were available at that time. Associated with these tools, there were already some techniques for predicting and inferring future positions based on analysis of past movements. These techniques were often used in order to navigate when, for instance, the sky was not clear and they were generally complemented by anchor points of known location, such as points on the shore. In the 18th century, the chronometer was invented. This tool, widely used in ship navigation, permitted calculations of position to be connected more efficiently to corresponding timestamps. By the end of the 19th century, wireless communications emerged and, along with this remarkable discovery, the first position solutions based on electromagnetic waves started to

[1] Or in space, if the system is ment to be used in outer space.

be developed. This period marked an important turn not only in the history of positioning systems, but also in the entire history of communication technology. The first positioning system to be invented was the Radio Direction Finder, a device which was able to determine the direction from where radio waves were being generated. The basic concept of this device was to find the null (i.e., the direction which results in the weakest signal) in the signal observed with a directional antenna mounted on a portable support. Only by the mid 20th century were the first radars invented. Ever since, these systems have been used and enhanced, and are still widely used for several positioning purposes. Radar is a system that transmits radio waves towards a target and is able to read the signals that are received back after they have been reflected from the target itself. From that period onwards, great development in positioning systems has occurred, culminating in the wide variety of systems currently available. One of the most famous systems was the Long Range Navigation (LORAN) system, proposed in 1940, which was characterized by beacons radiating synchronized signals that were then read by target receivers. The receivers had to be able to measure time differences of arrival of the signals in order to calculate their positions. Later on, by 1960 and 1970, satellite positioning systems started to be deployed, along with the beginning of space exploration. Currently, GPS is the best known and most used satellite positioning system. Over the past 30 to 40 years, several other positioning solutions have been deployed for use in various scenarios, based on various approaches, using infrared, ultrasound, image processing or electromagnetic waves.

In parallel with the deployment of positioning and communication systems during the last decade, research has evolved in the direction of integrating communication and positioning into a single system. Some examples are the current standards for the Third Generation Partnership Project (3GPP) and ultra-wideband (UWB), which include information about positioning mechanisms in the communication specifications. The combination of these functionalities has leveraged new services, namely LBS. In contrast to the earlier dedicated solutions that were designed to simply provide positioning, the new solutions for wireless networks are able to provide the combined benefit of both communication and positioning. As a consequence, the whole network, as well as the end user and the service providers, can benefit from position-enabled communication capabilities: the network operator can manage all the network's resources in a more efficient way, the service provider is able to deliver new services based on the user's position, and the user can enjoy new personalized, location-dependent services. Many of the services proposed in research documents assume location information as a basic requirement for deployment of new protocols (e.g., routing and clustering), new technologies (e.g., cooperative *systems*) and new applications (e.g., navigation and location-aware advertising). Furthermore, in the *industrial environment*, location information has also been widely used in applications such as navigation, location-dependent searching and social networks. Since wireless communication networks are nowadays almost present anywhere at any time, any

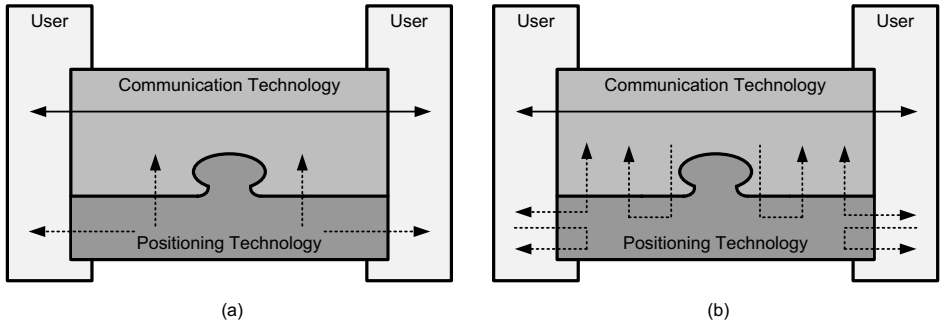


Figure 1.1 Schematic representation of a positioning system (a) and a positioning solution (b). The solid arrows represent the flow of communication and the dashed arrows represent the flow of position information. In (a), it is possible to see that position is always generated in a positioning technology. In contrast, a positioning solution is often calculated based on opportunistic information obtained from communication technology.

new location-dependent networking enhancement, service or application can be spread rapidly and used globally.

The location information mechanisms in wireless communication technologies can be classified into positioning systems and positioning solutions as in Figure 1.1. A positioning system concerns the hardware and software necessary to measure the properties of wireless links and to subsequently process those measurements in order to estimate the user's position. Positioning systems are typically integrations of entire systems into communication technologies, such as the GPS integration into mobile phones and cellular base stations (BSs). In contrast, a positioning solution is typically a software-based implementation, where indirect measurements of the user's position are obtained in an opportunistic fashion from the adaptation of mechanisms existing in communication technologies. The enablement of location information in current wireless communication networks can be done either by integrating a positioning system into the network or by implementing a positioning solution that extracts location information, thus exploiting the potential of the network. By integrating a positioning system into the network, it is usually possible to obtain better accuracy than what is obtained by a direct implementation in the network. The disadvantage is that integrating additional hardware necessarily implies additional costs, higher power consumption and higher complexity. For this reason, a direct software-based positioning implementation that opportunistically exploits the existing hardware and extracts location information is commonly preferred. Typical positioning solutions make use of mechanisms present in most of the current advanced wireless technologies such as power control and synchronization schemes. The requirement is that the positioning solution is able to perform measurements of the wireless channel that can be used as indirect observations of the user's position.

Examples of such measurements are the received signal strength and the propagation delay of signals.

The possible scenarios and conditions where location information is needed are numerous, and no localization system is able to perform localization under all conditions. For this reason, several localization systems and solutions are available and many others are still under research. For instance, the traditional satellite-based GPS represents a solution with worldwide coverage; however, it is not usable in some specific situations, such as in dense urban areas, indoors, underwater and underground. For these cases, many other systems are being developed or are even already available on the market.

The book is organized as follows. Chapter 2 introduces location-based services as the main driver for wireless positioning. Then, Chapter 3 introduces the basics of wireless communications. The foundations of positioning are presented in Chapter 4 and are then extended to more complex data-processing algorithms in Chapter 5. Tracking algorithms are explained in Chapter 6. As none of the previous chapters reviews the actual preprocessing of raw data, Chapter 7 introduces some algorithms for mitigating propagation effects in wireless positioning. The systems and solutions for wireless positioning currently existing in the market are described in Chapter 8, and, finally, some novel cooperative schemes for wireless positioning are introduced in Chapter 9. The sections below will give some insight of each of the chapters.

1.1 Application Areas of Positioning (Chapter 2)

Location-based services are introduced in this chapter as a motivational engine for combining communication and positioning information and then integrating wireless systems and positioning technologies. Owing to the increasing number of services and systems using location and the even larger number of services envisioned for the future, it has become mandatory to have a framework for layering positioning systems in a similar way to what has been done for the Open System Interconnection (OSI) stack in the communication world. This framework defines several layers, each of which has clear, defined purposes. This framework, apart from other benefits, will allow services to run independently of the overlaid technology. This chapter concerns mostly the applications of LBS and the entities involved in the services. These entities are, for instance, service providers, users, network operators, location infrastructure operators, developers and portals. The applications are categorized according to the service that is being provided: emergency services, automotive applications, medical applications, monitoring, navigation, management and entertainment. Then, the major benefits of combining the two functionalities are pointed out: interoperability and enhancement of both positioning-based communication and communication-based positioning. Typical examples are radio resource management and radio planning.

1.2 Basics of Wireless Communications for Positioning (Chapter 3)

In this chapter, the basics of wireless communications and how they relate to mobile positioning are introduced. First, the radio propagation scenario is described by classifying the various physical phenomena that underlie wireless communications, i.e., propagation loss, shadowing and multipath fading. Then, the possibilities and challenges of multiple-antenna techniques are discussed, focusing in particular on the two big families of *spatial diversity* and *spatial multiplexing* and on the gains achievable with them. The other important family of multiantenna techniques, i.e., beamforming, is discussed later in the chapter, when we will treat space division multiple access (SDMA), an access technique that is based on beamforming. The chapter reviews modulation and access techniques, focusing in particular on OFDM and orthogonal frequency division multiple access (OFDMA), which are the most relevant techniques for future wireless positioning systems. Moving up in the OSI stack, some radio resource management (RRM) techniques are illustrated, namely handoff, interference management, channel reuse, channel reservation and power control, and we illustrate how localization can benefit from them, and vice versa. Finally, we outlined some important ideas about two emerging areas within wireless communications, i.e., cooperative communications and cognitive radio technology. Their interconnection with mobile positioning closes the chapter.

1.3 Fundamentals of Positioning (Chapter 4)

After reviewing some basics of wireless communications in the preceding chapter, we introduce the basics of wireless positioning. Systems are classified according to several aspects: topology, physical coverage and type of integration into communication technologies. Then, the various types of measurements are explained. The cell ID is identified as the simplest type of measurement to obtain and also the one that results in the least accurate solution. Range measurements such as time of arrival (TOA) and received signal strength (RSS) measurements are introduced as having, apart from the effect of propagation impairment, the potential to provide more accurate positioning information than what is possible with Cell ID measurements. Two alternatives are measurements of the time difference of arrival (TDOA) and the angle of arrival (AOA). These measurements enable several basic positioning techniques that are explained in detail. Proximity-sensing techniques rely on measurements of proximity or coverage. Triangulation techniques rely on trigonometric theory and channel measurements in order to obtain the positions of mobile stations (MSs). Finally, fingerprinting techniques require an extensive precalibration phase that is aimed at collecting measurements of the wireless channels at several predefined positions. Then, while in operation mode, the measurements obtained are correlated with information from the calibration in order to localize the MSs. As none of the positioning systems are error-free, due to channel noise, the

chapter also introduces several of the most important types of errors existing in the measurements. These errors are mostly due to propagation effects, geometry and the technology used, and the consequent topology and communication protocols. Owing to the errors, it is necessary to define metrics of positioning. The most common metrics are the circular error probability (CEP), the dilution of precision (DOP) and the Cramér–Rao lower bound (CRLB).

1.4 Data Fusion and Filtering Techniques (Chapter 5)

Due to the various error sources, it is necessary to implement mechanisms capable of handling this noise. The first approach explained in this chapter concerns the least-squares (LS) method. This approach aims at determining the position that minimizes the squared error between the actual measurements from the channel and the expected measurements. In order to execute the algorithm, it is necessary to define a model to approximate the relation between the position coordinates and the measurements. In this context, linear and nonlinear methods are explained. While the linear LS methods present a closed-form solution and a derived method for recursive calculations, the nonlinear LS (NLLS) methods must be solved by numerical optimization algorithms. An alternative to the LS methods is the Bayesian framework. This framework defines a recursive mechanism to calculate the position based on the least mean squared error. Kalman filters (linear and nonlinear), particle filter methods and grid-based methods are illustrated as implementations of the Bayesian framework, each to be used depending on the properties of the system. As commonly happens in wireless positioning, there may be several parameters that are unknown, even though the model relating the measurements and the positions is known or at least can be approximated. For this reason, these parameters are determined by means of either precalibration or on line combined estimation of parameters and positions. The chapter closes by describing some alternative solutions, such as fingerprinting and time series analysis.

1.5 Fundamentals of Tracking (Chapter 6)

As a first step towards tracking of MSs, this chapter starts by conceptually explaining the complexity introduced into systems when time is considered, i.e., when users move. Then, as a further step in complexity, cooperative schemes are also introduced. Mobility models are presented as an essential aspect of tracking applications. The first class of movement models is that of the conventional models, which are given by the well-known physical laws of the straight-line movement. Then, several stochastic models for individual mobility and group mobility are presented. While individual mobility models tend to permit unconstrained movement or are constrained only by scenarios, boundaries, group mobility models tend to be constrained by the mobility of other individuals belonging to the same group. The final class of mobility models is that of cooperative or social based models, which resemble not only group

behavior, but also the social behavior of individuals. The latter models provide some of the foundations of cooperative wireless positioning presented in Chapter 9. After presenting the models, the chapter presents a sequence of components commonly present in a tracking solution. The first phase is to treat the raw measurements in order to exclude some possible destructive effects that can be detected at this early stage. Then, the measurements are used as an input for tracking algorithms; we first present non-maneuverable and then maneuverable algorithms. Finally, at the output of the tracking algorithm, it can exist an intelligent platform capable of learning users' positions and tracking patterns; two important algorithms are presented in this context.

1.6 Error Mitigation Techniques (Chapter 7)

Non-line-of-sight (NLOS) situations are commonly encountered in modern wireless communication systems in both indoor and outdoor environments. They result in biased time-of-flight estimates, which, if not handled properly in the localization algorithm, may considerably degrade the localization accuracy. In this chapter, common techniques that can be employed for the mitigation of NLOS effects in a wireless positioning system are reviewed for time-based positioning systems. After providing a generic system model for time-based location estimation, we summarize fundamental lower bounds and maximum-likelihood (ML) solutions for line-of-sight (LOS) and NLOS scenarios. Then, various NLOS mitigation methods are classified into four different categories and reviewed: (1) least-squares estimators; (2) constraint-based estimators; (3) robust estimators; and (4) the identify-and-discard-type of estimators. Least-squares algorithms improve the localization accuracy by giving less emphasis to NLOS measurements in the mobile location's solution. Constrained-based estimators use NLOS measurements to define certain constraints on the location estimate; under these constraints, they typically evaluate the location of the mobile station by using LOS measurements. Robust estimators make use of the rich body of literature and methods available in robust estimation theory for detecting and discarding outliers in a given set of data (i.e., NLOS measurements are treated as outliers and discarded). Finally, the identify-and-discard type of estimators aim at accurately identifying NLOS measurements and determining the location of the mobile station by using only LOS measurements. The key contributions in the literature associated with each of the above categories are summarized and compared with each other.

1.7 Positioning Systems and Technologies (Chapter 8)

This chapter presents in a concise way the positioning solutions commonly used in the various wireless communication technologies. The first class of systems to be introduced is the satellite-based systems. The basic principle relies on measuring the propagation delays that signals are subject to when traveling from the satellites to

the users on the Earth. Then, cellular positioning solutions are introduced; mainly considering Global System for Mobile Communications (GSM) and Universal Mobile Telecommunications System (UMTS) communication systems. While GSM typically relies on methods based on cell identification and range measurements, UMTS relies on differences of range measurements. The next class of systems introduced in this chapter is wireless local area network (WLAN) systems. The latter tend to use both cell identification techniques and measurements of received signal strength (RSS). The UWB technology is a special case, where time measurements are preferred, as such a wide frequency band permits time measurements to reach a higher accuracy than what is permitted by RSS measurements. Then, some dedicated solutions are presented: radio frequency identification (RFID), infrared and ultrasound. The chapter closes by summarizing some hybrid techniques that combine more than one technology in order to achieve better results in terms of position estimation accuracy. The combination of cellular and WLAN positioning with the Assisted Global Navigation Satellite System (A-GNSS), e.g., the Assisted Global Positioning System (A-GPS), has become of great interest.

1.8 Cooperative Mobile Positioning (Chapter 9)

Based on the knowledge acquired in the previous chapters, this chapter introduces cooperative schemes for wireless positioning. As a concept borrowed from the context of robotics, these schemes are described within that context and examined further within the context of sensor networks. As a natural requirement of cooperative schemes, nodes are required to cluster so that cooperative links are known within the cluster members and also within the entire network. With this in mind, wireless mobile networks are considered from the cooperative point of view. The estimation algorithms, previously introduced as based on individual positioning methods, in particular the least-squares and Kalman filter approaches, are augmented in order to consider several users and the influence between them. These algorithms are used as the core estimation process in the Cooperative Mobile Positioning System (COMET) framework, a cooperative infrastructure which provides location information on a group of users by taking into account the user-to-user communications and the proximity relations among them. The architecture is based on the concept of clustering the users, and then gathering the measurements from all the links between users and between users and BSs in order to jointly estimate the positions. The various mechanisms and the entire COMET concept are analyzed based on simulations.

2

Application Areas of Positioning

2.1 Introduction

In this chapter, we start by presenting the Location Stack, a framework created by Jeffrey Hightower (Hightower et al. 2002), which offers an overview of an exemplary localization system with all its elements and interrelations, similar to the seven-layer ISO/OSI model used in the field of communication systems. Then, we continue by introducing the first application area of positioning and thereby the formal definitions of a location-based service (LBS), which is basically considered as the intersection of new information and communication technologies (NICTs), the Internet and geographic information systems (GISs). Further, we illustrate the "LBS ecosystem", which is an efficient parallel taken from biology, to show the many actors rotating around this business; in particular, such actors will be distinguished into *operational* and *non-operational* actors. The chapter follows this by presenting taxonomies of location-based services, which are divided into the *service model*, *application*, *market segment* and *content*. We shall focus on the second of these classes and define ten application categories within location-based services. Finally, we shall proceed to the second application area of positioning, where location information is exploited to enhance network performance rather than to provide new services to the user. In this context, we shall consider radio network planning and RRM issues closely.

The rest of the chapter is organized as follows: Section 2.2 presents the Location Stack, Section 2.3 introduces location-based services and Section 2.4 describes location-based network optimization issues. Finally, some concluding remarks are made in Section 2.5.

2.2 Localization Framework

Before we describe the wide set of applications related to positioning, it is necessary to give a general picture of the various elements involved in a localization system.

© 2010 John Wiley & Sons, Ltd

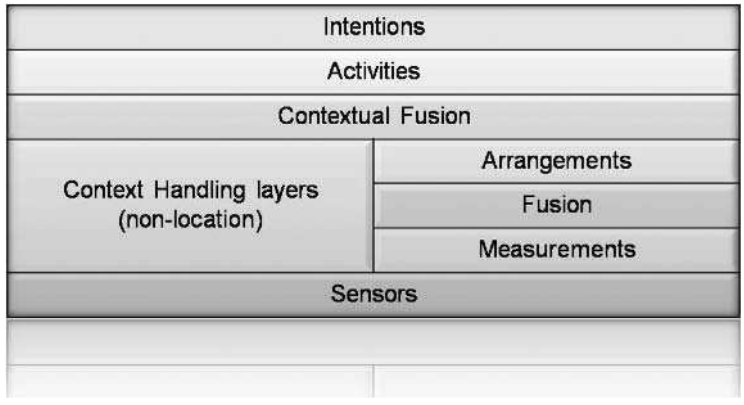


Figure 2.1 The Location Stack (Hightower et al. 2002).

Hightower et al. (2002), by means of a survey of the literature, were able to extrapolate five design rules and make an abstraction of a localization system; practically, this resulted in a seven-layer model similar to the ISO/OSI model for communication systems (Figure 2.1).

The contents and actions of each of the layers in the seven-layer model for localization systems can be briefly described as follows:

- **Sensors** (Layer 1) represents the available sensors in a localization system (e.g., infrared badges, barcode scanners, cameras, pressure mats, thumbprint readers and keyboard login systems) and provides as output a stream of raw data (e.g., IDs, blob pixels, binary states and login events).

- **Measurements** (Layer 2) represents the algorithms that turn this data flow into its canonical form (e.g., proximities, distances and angles). Along with that, it also provides a model of the uncertainty that is characteristic of the sensor that generated the information.

- **Fusion** (Layer 3) represents the data fusion algorithms (see Chapter 5), which combine the retrieved data from all of the available sensors to calculate the coordinates of the target objects or persons (see Chapter 4).

- **Arrangements** (Layer 4) represents the algorithms that interrelate the positions of these targets, for example by converting their coordinates according to a relative coordinate system (e.g., from the position of a person in a room to the position of that person with respect to the loudspeakers in that room).

- **Contextual fusion** (Layer 5) represents the system that merges the pure location information with other contextual information (e.g., information ranging from personal data to temperature and light level).

- **Activities** (Layer 6) represents the system that, thanks to the contextual information retrieved in Layer 5, recognizes the current activities of the targets (e.g., whether all family members are asleep).
- **Intentions** (Layer 7) represents the system that, based on current activities and predefined user plans, takes decisions about the actions to be undertaken (e.g., increasing or decreasing the temperature in the bedrooms by a few degrees).

While Layers 1–4 will be mostly covered in later chapters, Layers 5–7, which can be seen as elements of the application layer, will be touched upon here.

2.3 Location-based Services

An LBS, in the broadest sense, is any service that extends spatial information processing or GIS capabilities to end users via the Internet and/or wireless networks (Koeppel 2002). However, unlike conventional GIS applications, which are focused on geographic data for land management and planning, LBS applications can provide both the connectivity and the context required to dynamically link an individual's location to context-sensitive information about the current surroundings. This enables a high level of personalization, which facilitates an ability to "put each user at the center of his/her own universe" (Kivera 2002). Some more formal definitions of an LBS are the following:

- **LBS definition 1.** A wireless IP service that uses geographic information to serve a mobile user or any application service that exploits the position of a mobile terminal (OGC 2003).
- **LBS definition 2.** An information service accessible with a mobile device through a mobile network, utilizing an ability to make use of the location of the mobile device itself (Virrantaus et al. 2001).
- **LBS definition 3.** A network-based service that integrates a derived estimate of a mobile device's location or position with other contingent information so as to provide added value to the user (Green et al. 2000).

The above definitions practically describe an LBS as an intersection of NICTs, the Internet and a GIS (Brimicombe 2002; Shiode et al. 2004) (Figure 2.2).

2.3.1 LBS ecosystem

Inspired by biology, we apply the metaphor of an "ecosystem" to refer to a location-based service as a whole composed of several different actors, their roles and their interacting synergism, which forms a complex web of interdependencies. As defined in Küpper (2005), an actor is "an autonomous entity (e.g., a person, a company or an organization), which adopts one or several roles that characterize either the functions it fulfills from a technical point of view or the impacts it exerts on LBSs from an

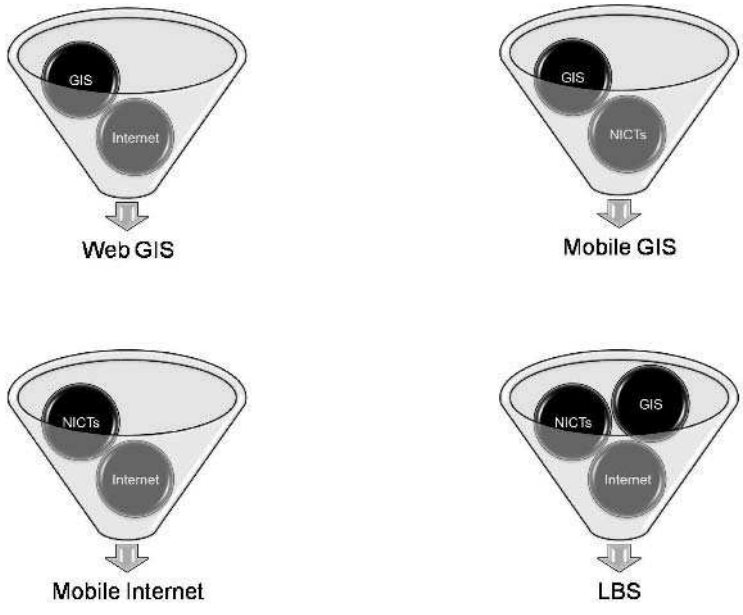


Figure 2.2 An LBS as an intersection of technologies.

economical or regulatory point of view". As a consequence of this definition, we can easily extrapolate to say that two main categories of actors exist: the *operational* actors, which are directly involved in the practical operation and functioning of LBSs, and the *nonoperational* actors, which have an indirect influence on the overall operation of LBSs. Figure 2.3 illustrates the components of the LBS ecosystem, whose roles are briefly described below (Küpper 2005; Steinfield 2004):

- **Operational actors:**
  - **End users:** these are the final recipients of location-based services.
  - **LBS providers:** these offer location-based services to end users by means of the network operators' infrastructure.
  - **Network operators,** i.e., mobile and wireless network operators and mobile virtual network operators (MVNOs): these own and manage the network infrastructure and the location infrastructure, from which they retrieve or collect the positioning data, and perform billing operations and provide services to third parties. Sometimes they also interface directly to end users by offering them LBSs, thus acting in such a context as LBS providers.
  - **Location infrastructure providers:** these sell mobile location centers and other hardware and software to network operators.

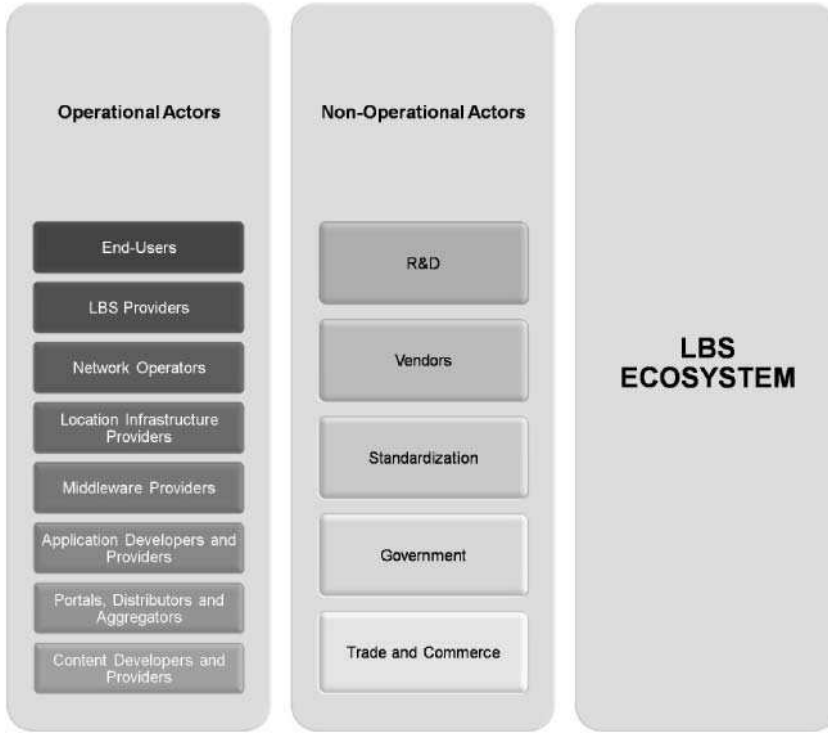


Figure 2.3 The LBS ecosystem.

  - **Middleware providers:** these offer tools that facilitate the use of the various applications made available by application providers to LBS providers.
  - **Application developers and providers:** these create applications that utilize the content gathered by portals, distributors and aggregators, and supply them to LBS providers.
  - **Portals, distributors and aggregators:** these gather the various content offered by the content providers and make it available to application developers.
  - **Content developers and providers:** these consist of GIS application service providers and producers of geographically referenced information – from the location of streets and buildings to points of interest and real-time information such as news, weather and traffic, and mapping services.

- **Nonoperational actors:**

  - **R&D:** this embraces universities, research centers, and companies with internal research and development units. This category continuously

brings forward and patents novel technological solutions, and thereby dictates the pace at which location-based services progress.

- **Vendors:** these include manufacturers of handsets and of positioning and network-related equipment. By offering tools to enable and interact with location-based services, this category has a clear impact on network operators and end users.
- **Standardization:** this encompasses the committees that take care of defining the design principles of location-based services in terms of interfaces, protocols and application programming interfaces (APIs); this process influences a range of actors in the operational chain, from application developers to LBS providers. Note that it is usually the presence of a standard that dictates the success of any given technology, since many proprietary solutions, which by definition do not share a common platform, would emerge otherwise; this makes communication between the various actors involved in the LBS ecosystem difficult and reduces the chance to provide end users with seamless, ubiquitous services.
- **Government:** this comprises the governmental units responsible for legislative issues related to location-based services. This category legislates on regulatory (nontechnical) matters – for example, the establishment of rules to be followed when location data is handled, by defining the boundaries between privacy and lawful interception – which affects operational actors such as network operators and LBS providers.
- **Trade and commerce:** this consists of the companies that make direct or indirect use of location-based services to empower their internal efficiency, their final products and their marketing strategies. On the basis of feedback from this category, developers are required to adjust their applications.

2.3.2 Taxonomies

Location-based services can be classified according to the following approaches (Montalvo et al. 2003):

- **Service model.** This approach categorizes a service according to (Nepper et al. 2008) (1) the direction of session initialization, (2) the relationship between user and target (i.e., the target may be the same user, other users, an object, etc.), (3) the number of targets and (4) the type of target. The first criterion results in a distinction between *reactive/pull services* and *proactive/push services* (Figure 2.4). While the former are triggered by the user, who reactively "pulls" location-based information from the network (e.g., by asking for a reasonably priced café nearby), the latter are triggered autonomously by the infrastructure, which proactively "pushes" location-based information towards the user, owing to the expiry of a timer or the occurrence of a geographically co-located event

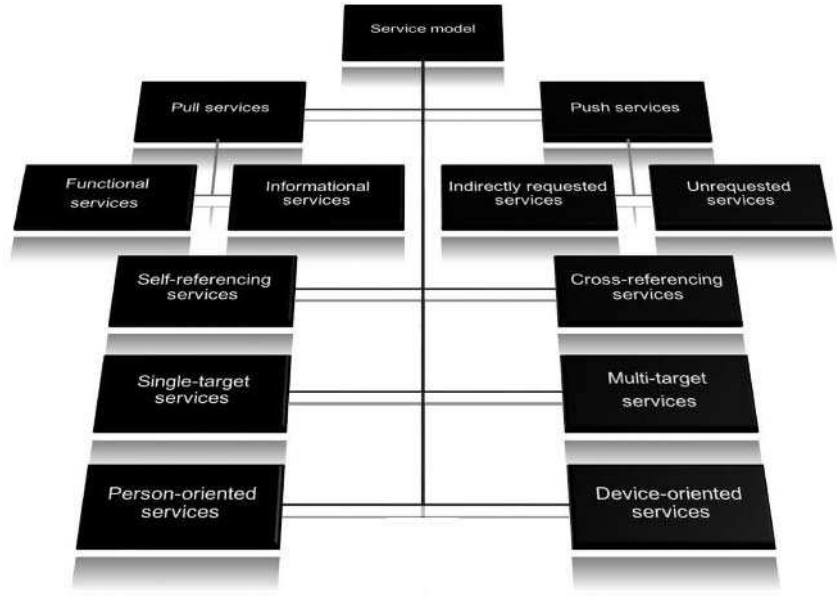


Figure 2.4 Service model.

(e.g., a user in the surroundings of a café which offers a 20% discount at lunchtime). Note that some services, such as the Friend Finder, embrace both pull and push functionalities (Schiller and Voisard 2004). A further separation can be done for pull services into (Virrantaus et al. 2001) *functional services* (e.g., ordering a taxi or an ambulance just by pressing a button on a device) and *informational services* (e.g., searching for a Chinese restaurant nearby). For push services, we can make a further separation into (Steininger et al. 2006) *indirectly requested services* (e.g., a local news service subscription) and *unrequested services* (e.g., a hurricane warning message). The second criterion results in a distinction between *self-referencing services*, where the user and the target are the same entity, and *cross-referencing services*, where the user and the target(s) are separate entities (e.g., a user who is looking for a date in their proximity). The third criterion results in a distinction between *single-target services* (e.g., where a user can visualize the position of a shopping mall on a map) and *multitarget services* (e.g., where a user can visualize on a map the positions of all users matching the first user's personal profile). Finally, the fourth criterion results in a distinction between *person-oriented services* and *device-oriented services* (Schiller and Voisard 2004). The first class comprises services that exploit the target user's position to enhance or enable certain applications, and are controlled by the users themselves (e.g., personal navigation). The second one, in contrast, embraces services that make

use of the target device's position for actions, that are external and thereby not controlled by the user of that device (e.g., tracking of criminals).

- **Application.** This approach categorizes a service according to its main target application. Specifically, we have derived ten categories (see Section 2.3.2.1): *emergency, safety and security*; *tracking and tracing*; *traffic telematics (including vehicle navigation)*; *personal navigation*; *management and logistics*; *billing*; *commerce*; *enquiry and information*; *leisure and entertainment (including community and gaming)*; and *supplementary*.
- **Market segment.** This approach categorizes a service according to its main target market (Niedzwiadek 2002), i.e., the consumer, business public sector market.
- **Content.** This approach categorizes a service according to the type of information provided (Van de Kar 2004): positions, events, distributions, assets, service points, routes, context, directories and transaction sites.

2.3.2.1 Application categories

The categorization below has been done by collecting together all of the different and even discordant inputs present in the literature and trying to make them converge as much as possible into a consistent view (Figure 2.5). In particular, the services listed in each category were not aimed at forming a definitive list, as new services are continuously emerging. Many of them are evidently related to more than one category. However, we have assigned each service to the category that it is most directly linked with. Finally, note that while one can find *infotainment services* (information services + entertainment services) as a stand-alone group in the literature under either the category "information" or "entertainment", we have decided to spread them between these two categories, specifically, between what we have identified as *enquiry and information* (Section 2.3.2.1.8) and *leisure and entertainment* (Section 2.3.2.1.9).

2.3.2.1.1 Emergency, safety and security[1] The most immediate application of location-based services is the enablement of emergency services, i.e., services that help users in critical situations, where their safety and security are under threat. Indeed, there are occasions on which we are not able to inform the emergency center of our location (e.g., when we have been involved in a serious traffic accident), we are not able to reveal it (e.g., if we are currently being subjected to a criminal or terrorist attack) or we simply ignore it (e.g., when our car suddenly breaks down). Note also that in many cases, under stressful or shocking situations, callers do

[1] S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part with kind permission of © 2007 Springer Science and Business Media.

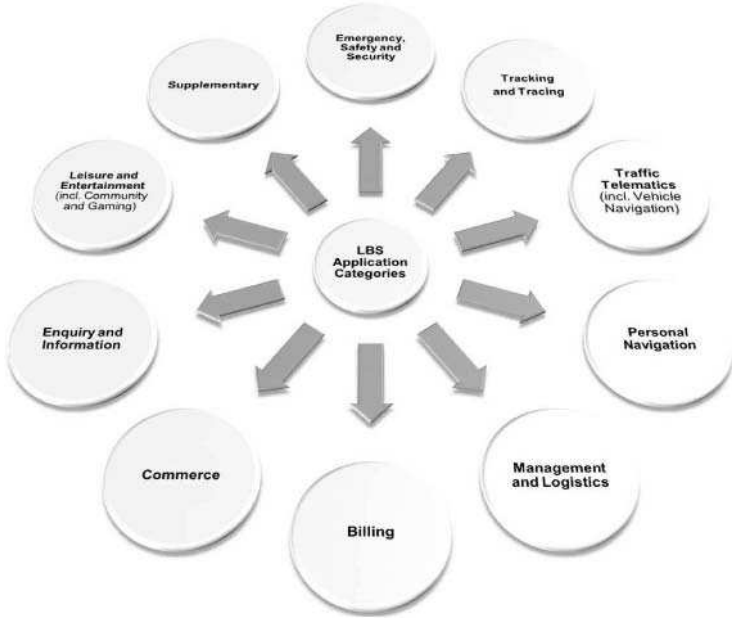


Figure 2.5 LBS application categories.

not always act calmly and provide the information strictly necessary to allow the service center to immediately dispatch the emergency units needed (e.g., police, firefighters or paramedics). In addition, if users/devices (e.g., mobile phone or car) are able to signal their location, the emergency can be treated much more efficiently, which translates into precious time saved (and consequently more lives saved), higher user satisfaction, etc. While emergency call services are usually regulated by public organizations and therefore are made available to mobile phone users without requiring any pre-subscription, other types of service, such as automotive assistance, are managed by private companies. Table 2.1 shows examples of emergency, safety and security services, which we have assessed in terms of accuracy, environment, business type, revenue potential and related categories.

Practically, these services have driven the development and deployment of positioning technologies. Indeed, geolocation has gained considerable attention over the past decade, especially since the Federal Communications Commission (FCC) in 1996 passed a mandate requiring cellular providers in the USA to generate accurate location estimates for E911 services (FCC 1999, 2001) (see Table 2.2[2]). A similar mandate was extended to the EU in 2003, where mobile positioning is considered an even more critical issue, owing to the continually increasing number of mobile-originated E112 calls (EUROPA 2003). Nevertheless, the EU will most likely follow

[2]For the definition of "CEP67", "CEP95", and network-based and mobile-based positioning, see Chapter 4.

Table 2.1 Examples of emergency, safety and security services. (Reproduced with kind permission of © 2007 Springer Science and Business Media)

Service	Accuracy[a]	Environment	Business type	Revenue potential	Brief description	Related categories
Emergency call	High	Indoor/outdoor[b]	Business to Consumer (B2C)	High	This service allows one to quickly respond to an emergency situation by dispatching police, medical personnel, etc. to the location at which the call was originated (TruePosition 2009).	Tracking and tracing
Automotive assistance	Medium	Outdoor	B2B/B2C	Medium to high	This service allows one to automatically send help directly to the location of a damaged or inoperable vehicle. Sophisticated location-enhanced automotive assistance services offer useful on-the-road services, such as providing distances and directions to gas stations, automobile repair services or motels (AAA 2009).	Tracking and tracing
Emergency alert	Low	Outdoor	B2C	Low	This service allows one to deliver emergency information such as information about inclement weather, road conditions, heavy traffic or any life-threatening incident that is relevant to the user's location (Telenity 2009).	Tracking and tracing
Personal medical alert	High	Indoor/outdoor	B2C	High	This service allows a person who requires medical attention to simply press a button on a device (or type a preset number into a mobile phone) to alert an emergency response service, so that an ambulance can be quickly sent to the appropriate location. This saves precious moments in the response time and provides location information even when the person is unable to speak. This service is ideal for senior citizens and subscribers with life-threatening health issues (TruePosition 2009).	Tracking and tracing
Family monitoring	High	Indoor/outdoor	B2C	Medium	This service allows one not only to locate or monitor a child or elder, but also to provide an alert message if the person moves outside a predetermined area; this is commonly referred to as *geo-fencing* (TruePosition 2009).	Tracking and tracing

[a] Accuracy: high, from 0 m to 50–75 m; medium, from 50–75 m to 250–300 m; low, from 250–300 m onward.

[b] "Outdoor" means urban, suburban and rural.

Table 2.2 FCC requirements

Specification	Network-based	Mobile-based
CEP67	FCC-N 1: 0.1 km	FCC-M 1: 0.05 km
CEP95	FCC-N 2: 0.3 km	FCC-M 2: 0.15 km

the USA and Japan in requiring high positioning accuracy from 2010, when Galileo will be fully operational (Berg-Insight 2006). In the meantime, research in the field of wireless location has been promoted as an important public safety feature, which can also add many other potential applications to future telecommunication systems (Sayed et al. 2005a).

2.3.2.1.2 Tracking and tracing This category embraces services that are able to track and trace the geographic whereabouts of mobile entities (e.g., people, vehicles or packages) and support requests to establish the location of these entities, their progress and state changes along a route, or to establish prospective future locations (e.g., tracking postal packages so that companies know where their goods are at any time) (Steininger et al. 2006). In particular, these services can be distinguished according to their tracking target, i.e., a single entity (e.g., a child or a package) or a group (e.g., a workforce); thus, it is naturally derivable that this category may be directed both to the consumer and to the corporate market. Indeed, applications include, on the one hand, pet, people and vehicle tracking, and, on the other hand, fleet, asset, goods and package tracking. Obviously, there is an intertwining between this category and other LBS categories, such as emergency, safety and security, and management and logistics. For example, vehicle tracking can be also applied to dispatch the nearest ambulance to a given emergency call. A similar application can help companies to manage their workforce (e.g., salespeople and repair engineers) more efficiently, so to provide their customers with accurate arrival times of personnel (Steininger et al. 2006).

2.3.2.1.3 Traffic telematics (including vehicle navigation) The area of traffic telematics is aimed at supporting drivers with a manifold set of services related to their vehicles: vehicle navigation, diagnosis of malfunctions, dissemination of location-based information and warning messages (e.g., traffic, weather and road conditions), automatic configuration of appliances, and added features in the vehicle (Küpper 2005). An emerging research topic within this category is intervehicle communications, which usually rely on short-range wireless technologies (e.g., ZigBee, Wi-Fi or Bluetooth) to establish ad hoc communications between two vehicles without the need for a centralized control. This allows users to interact directly with each other and exchange useful information such as warning messages about accidents, local traffic situations or the position of filling stations (Küpper 2005).

The most popular application in traffic telematics is undebatably vehicle navigation. Whereas for the last decade this service has been enabled by expensive on-board units installed in vehicles, recently we have witnessed the growing integration of GPS receivers in portable devices, which can easily be plugged into our cars. Basically, whatever the device in use, we just need to set the location of our final destination and our terminal is able to locate our current position, calculate the route based on our preferences (e.g., the shortest, fastest, most economical or most scenic route) and feed this information back to us in the most suitable way (i.e., according to the capabilities of our device and our preferences). In fact, this information can be delivered in text format, on a map or with the help of automatic voice guidance, or by a combination of these. If our terminal is able to connect to a cellular network, we are also able to receive up-to-date traffic telematics information related to our vehicle navigation (e.g., traffic jams and road works); on the basis of this information, our device can also recommend alternative routes. Finally, vehicle navigation can be combined with a number of services related to parking slots, ranging from registering and charging to automatic guidance and exchange of parking slots among drivers (Küpper 2005).

2.3.2.1.4 Personal navigation While traffic telematics embraces vehicle navigation, personal navigation is in its own category. Specifically, this category has emerged from the increasing pervasiveness of mobile devices in our lives (from personal digital assistants (PDAs) to mobile phones) and their continuous and progressive enrichment of integrated technologies (from various wireless communication to positioning technologies); in this context, mobile positioning will soon be enabled in both outdoor and indoor environments. Indeed, users are not anymore concerned about their navigation only while driving, but also while walking, which certainly also includes indoor environments (e.g., shopping malls and offices). Practically, one main difference with respect to traditional vehicle navigation is that users get a precalculated route via their mobile network operator, as road databases are in many cases not available on the device. Moreover, being not restricted to roads, personal navigation may possibly feature positioning algorithms different from those employed in vehicle navigation, and the visualization of maps may also be different.

2.3.2.1.5 Management and logistics The services considered under this category encompass any type of management, from facility to environmental management. The inclusion in such management processes of "location intelligence" is a key asset for enabling higher efficiency and reducing possible risks. For instance, in the case of a complex construction project, such smartness allows easy management of work crews as they move from location to location. As outlined in (Graphisoft 2007), "Usual scheduling systems are activity-based (good enough for building an airplane engine in a factory), which implies that the user manually takes into account locations and makes subjective determinations on the duration of a certain activity. The addition of location as an extra dimension provides enormous benefits, since it allows users

to compress schedules without increasing risks by balancing the amount of float throughout the project and, therefore, reducing workflow variability." However, there are still open issues that may reduce or delay the provision of such benefits. In the case of facility management, for example, inspectors often find it hard to acquire and manipulate location data about facilities when the facilities do not have any distinctive location marks to be referenced. Although they are equipped with GPS-enabled devices, this may still happen if the identification of the location of the facility is housed inside a building (Tejavanija et al. 2003). Similar considerations and advantages, deriving from the opportunity to locate and track a workforce, apply also to the case of fleets (e.g., freight, public transportation and emergency vehicles). For instance, the operator of a taxi dispatch center may have all the positions of their company's vehicles displayed on a map. When incoming customers need to be served, the operator just informs the target taxis (possibly the closest to the callers) of their new destinations, and the users of the expected arrival times. Finally, it is important to mention that knowing the location of a product and being able to track it can also serve to support logistics. Indeed, this makes it possible to support faster transportation, different transportation modes and the development of fallback scenarios in the case of failures (Küpper 2005).

2.3.2.1.6 Billing This category can, in one instance, be related to mobile users, who can be charged by their wireless carriers on the basis of their current location. Forms of facilitation and service differentiation such as predefining specific zones (e.g., "home zone", "work zone" and "premium price zone") with predetermined rates can be set up (e.g., calls that originate within the user's home zone may be charged as wired calls). In this context, it is also possible to define areas in which callers are allowed to reach the customer and areas in which calls must be blocked. For example, the user may want to receive certain calls in the home zone, but not at work, and perhaps receive only urgent calls when traveling or on vacation (MobileIN 2009). Finally, we can also expect that operators, in order to maximize their revenue, will offer special types of services in agreement with other service providers. For example, as Kenney Jacob foresees, we can think about an advertisement in front of a Coffee Day shop, which says (Jacob 2007) "Enjoy free Hutch to Hutch Calls from your mobile along with your Coffee". This obviously also reminds us of the fact that location-sensitive billing is often found in combination with the purchasing of goods and services, and therefore is an enabling feature of mobile commerce applications (see Section 2.3.2.1.7).

Billing, i.e., charging services in connection with the user's vehicle, can also be related to traffic telematics. Road tolling, which concerns the payment of tolls related to the use of roads, highways, tunnels, bridges, etc., is the most evident example of this type. In general, such payments have not yet been completely automated; indeed, most of the time, we stop at a barrier and pay our toll to an officer or a machine according to the distance covered. Since such an approach increases traffic congestion, some countries have introduced a subscription, which is valid for a certain

time period and permits the driver to circulate freely. However, even this solution has the shortcoming that drivers are not billed for the distance that they have really covered (Küpper 2005). Currently, the only alternatives for avoiding such scenarios and enforcing payment automatically are the following: (1) vehicles carry an on-board unit, which, thanks to a transponder, is able to exchange data over the air with the toll barrier; and (2) control stations take pictures of passing vehicles and electronically analyze the license plate number by image recognition. Unfortunately, each country in the past has developed and deployed its own proprietary system and, therefore, incompatibility issues have emerged when national boundaries are crossed. This is one of the issues that has to be tackled within a framework of globalization, as there is a strong need to create international efforts to harmonize such a fragmented picture. Finally, it is important to underline that road tolling is only one of many services that fall under the broader umbrella of *remote tolling*. The latter services can be listed as follows (Antonini et al. 2004): (1) pay-per-use services (e.g., automotive assistance); (2) location-based information services (e.g., traffic information and routes); (3) tracing, tracking and emergency management services for the transportation of hazardous material (HazMat) and heavy goods; (4) access control services (e.g., enforcement of laws against users who enter restricted zones such as historic areas without permission); (5) parking services (see Section 2.3.2.1.3); and (6) pay-per-use insurance: users pay for the insurance of their vehicles on the basis of the percentage of time the vehicle is used, the type of environment in which the driver normally uses the vehicle, and on-board information that can affect the final insurance cost.

2.3.2.1.7 Commerce In this category we consider all applications related to commerce, i.e., from marketing to the purchasing of goods and services. Obviously, location information opens up new sale scenarios and provides marketers, and thereby companies, with great potential to increase their revenue by intercepting new customers and potential to empower their customer relationships by providing higher customer satisfaction. This can be achieved by means of targeted advertisements, which are activated and directed to potential buyers according to the buyer's personal profile, buying pattern in the past and proximity to the company's premises. For instance, users passing a shop could be attracted inside by an offer popping up on their mobile phone, where this message could refer to discounts, coupons, allowances and gifts, and could be presented in the form of full-color images, short videos or audio files. For larger stores, such as superstores and hyperstores, customers' navigation through the store should also be supported (Ad2Hand 2009). Naturally, this type of advertisement is applicable not only to users walking by, but also to people is traveling by public transport or driving. However, in the latter case, it must be guaranteed for safety reasons that the user will not be distracted by such messages (Küpper 2005). A further point of meditation regarding such services is related to the avoidance of spam. In this respect, Donald Spector, Chairman of the Innovation Fund LLC, says (IF 2009): "If you call someone and offer them a 10,000 dollars

promotional bonus from GM while they are standing in a Honda showroom, they will probably not get mad. We are not going to be bugging people with 25 cent offers. Shopping malls, restaurants, entertainment venues and the travel industry will be able to give consumers great bargains on items that otherwise would lose value. Tickets of a baseball game that is not sold out can be directed to a passing car making the customer an 'offer he can't refuse' ". "Remember," he adds, "we already know who likes baseball!". From this we can derive how important it is to filter the information according to the users' preferences, such that no one will be "bugged" by a massive load of unsolicited messages, which in the long run would lead to a rapidly decaying interest of the customer in this type of services. As a consequence, it must also be possible to cancel a subscription either permanently or temporarily.

2.3.2.1.8 Enquiry and information These services target the distribution of all of the pieces of information that are relevant to the user's context. This is, so far, the most widespread type of LBS and can embrace both users who live in a certain area and users who are there for the first time (e.g., for business or leisure purposes). Location-sensitive information concerns guided tours, transportation services and points of interest (e.g., nearby places of historical and touristic interest, hotels, restaurants, coffee shops, shopping malls, cinemas, pubs, discos, theaters, museums, parks, pharmacies, hospitals, gas stations, ATM counters and retailers). Services that provide points of interest are practically an extension of the well-known Yellow Pages that show entries only of local relevance. In particular, they may be combined with navigation facilities for guiding the user to a point of interest or even inside a point of interest (e.g., in a guided tour of a museum) (Schiller and Voisard 2004). Finally, we also recognize under this category services related to the distribution of information closer to the realm of traffic telematics, such as information about traffic jams, parking slots, and road and weather conditions.

2.3.2.1.9 Leisure and entertainment (including community and gaming) Community services underlie the gathering of people in either a physical or a virtual place around some common baseline, which might relate to keeping in touch with family and friends, reactivating connections with old friends and acquaintances, or creating discussion groups on specific topics (from cooking to traveling to eroticism). Usually, each user creates their own profile (including their alias, sex, age, domicile, hobbies, work, marital status, etc.), which is used for discovering people with the same background, interests and hobbies, and communication among the members of the community is enabled by means of chatrooms, whiteboards, etc. Recently, there has been an explosion of such services – see Figure 2.6 (Ziv and Mulloth 2006) – thanks to the increasing and widespread use of Internet, which consequently has facilitated a tendency towards meeting in "virtual places" and therefore creating new social links or empowering existing links regardless of the real physical distances between people. The inclusion of location information in such community services has launched a new paradigm: from people to people irrespective of

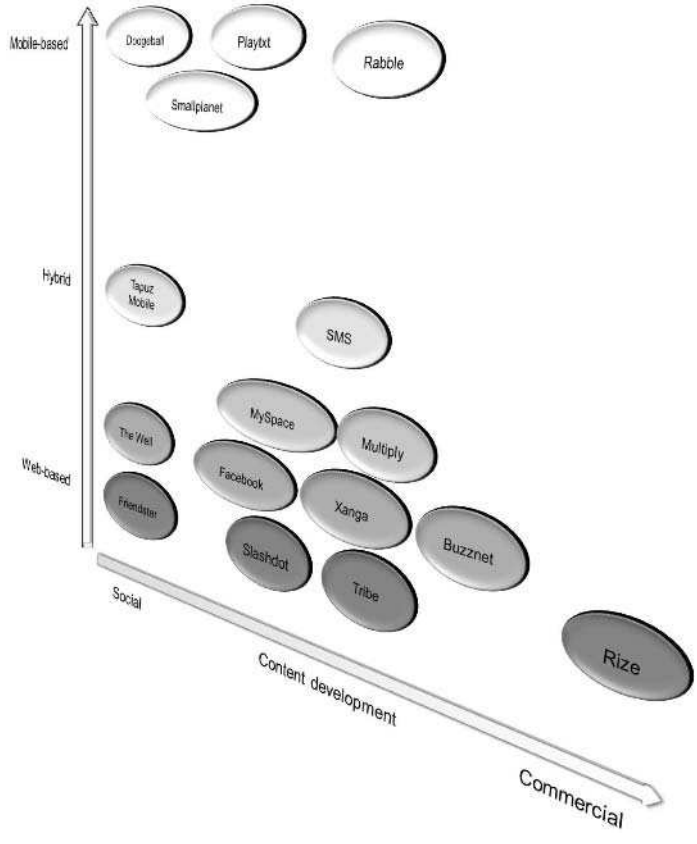


Figure 2.6 The social network matrix (partial list) (Ziv and Mulloth 2006).

geographical place to "people-to-people-in-geographical-places" (Jones and Grandhi 2005). Dodgeball, for example, is one of these services, where location information is used to meet nearby friends, friends of friends or even possible crushes (Dodgeball 2009). In the context of community services, privacy is a highly sensitive matter, as the level of information that we are willing to make accessible to others depends strictly on the level of social connection that we have with them. That is why, on the one hand, it is important to enable "permission request" messages for every enquiry we receive and, on the other hand, to explicitly declare in our profile the level of detail at which we want to give out information regarding our location. For instance, we can expect that for services such as a family finder or friend finder, the exact position of the user may be freely accessible to both parties, whereas for proximity-matching services or dating services, releasing the location of the user within a certain "privacy range" would be more suitable.

As a primary form of entertainment, at least for youngsters, gaming has evolved tremendously in recent decades, where the trend has moved from a single-player

or multiplayer game on a single PC to real-time multiplayer sessions where several computers are connected together in a local area network and to massive multiplayer online games with hundreds or thousands of people connected to the Internet. Recently, owing to the introduction of powerful consoles such as the Play Station Portable (PSP), gaming has gone portable. The next advance in this field will be the introduction of location awareness, which allows users to play in a real environment instead of a virtual one. For example, a shop that is willing to promote its products may sponsor a treasure hunt in a specific location, where the "treasure" is a 29 inch flat-screen TV. The first player reaching that location can claim the prize. Finally, it is important to mention that the next level is foreseen to be the introduction of *augmented reality*, which will allow users to mix reality and virtuality into the game. This technology will further blur the line between what is real and what is computer-generated by enhancing what we see, hear, feel and smell (Schiller and Voisard 2004). This could also create new prospects for worlds that are currently totally virtual, such as Second Life (Second Life 2009).

2.3.2.1.10 Supplementary Usually, "value-added services" refer to services which enrich the panorama of the user experience beyond basic speech or data services. In practice, many of the services presented above, mostly in the category "enquiry and information", can be included in this category. That is why we prefer to entitle this section "Supplementary", so that we could include all services that fall outside of the previous categories. Practically, the main supplementary services, which can be extended to location-based services, are the following. (1) Call forwarding (or selective routing): this service is usually activated when incoming calls addressed to a user's mobile phone are derouted to a co-located fixed telephone. (2) Coverage information intimation: this service can alert users when they are moving out of coverage or out of a service area (e.g., "high-speed data services end after 2 km") (Chaudhari 2007). (3) Voting: this service can gather votes (e.g., in relation to politics, various events or products), which may have strong local significance. If a regional political election is approaching, for example, political parties can obtain strategic surveys by gathering electronic votes from the very people on the spot, people who are really informed about the local problems.

2.4 Location-based Network Optimization

In contrast to location-based services, where the user is the main recipient (either passive or active) of such services and the target is mainly to provide appealing applications that benefit from the notion of location, here we are concerned with the optimization of some network functions by making use of this additional information. In practice, this process is transparent to the user and targets mainly the quality-of-service (QoS) guarantee provided by the network. In particular, while current mechanisms for network optimization are usually reactive, i.e., they counteract

current, short-term situations with the help of analysis of measurements, location-based intelligence permits the operator in most cases to make proactive decisions, which help to prevent the network suffering from poor performance.

2.4.1 Radio network planning

Radio network planning is a process that has the scope to provide ubiquitous coverage to mobile users while conforming to their expected QoS requirements (e.g., a low percentage of dropped calls or good indoor coverage). This process evolves in the following three phases (Chaudhari 2007): *dimensioning*, *detailed planning* and *optimization*. Starting from specifications of coverage, capacity, QoS demands and target service areas (e.g., higher and lower-data-rate services), and plausible radio propagation models, the first phase is characterized by an initial shaping of the network itself, which concerns the number of base stations (BSs) or access points (APs), their sites and their configurations. The second phase concerns more advanced planning of coverage, capacity and frequencies based on an analysis of real-time propagation data (usually gathered during a walk or a drive test) and then on the use of planning tools (see, for example, SEAMCAT at www.ero.dk), which permits one to check for interference and to define frequency reuse patterns, antenna settings, etc. Finally, the third phase consists in a continuous optimization process, which is active during the operation of the network and allows one to dynamically refine the configuration of the parameters initially set during the dimensioning and detailed-planning phases (e.g., transmission powers). In this context, it is obvious how location data can be very useful for identifying possible problems in the network (e.g., areas of poor coverage), for tuning and validating propagation models with real-time data and for creating alternative network plans in the case of traffic congestion (CELLO 2003). During sporting events, concerts or exhibitions, for example, cellular operators could insert additional capacity by adjusting the antenna settings or by using movable BSs (Lathi 2000). In general, location intelligence is an additional tool that can also be exploited prior to radio network planning to perform accurate market and demographic analyses, which will give indications about where and how to expand the network in order to get the maximum return from the initial operator investment (MapInfo 2009).

2.4.2 Radio resource management

The aim of radio resource management (RRM) is to utilize the available, often rather limited, radio resources as efficiently as possible. Here, "efficiency" refers to maximizing the traffic load while satisfying the QoS requirements of as many mobile users as possible and overcoming the inherent difficulties of radio wave propagation in the environment (Hasu 2007). This section provides an overview of several RRM techniques (for details, refer to Chapter 3) which can be optimized by introducing the notion of location-awareness.

2.4.2.1 Beamforming

Beamforming is a signal-processing technique that permits one to control the directionality of a radiation pattern in order to increase the sensitivity of a transmitting/receiving antenna array in the direction of a desired signal while decreasing it in a direction of interference and noise. Basically, this result is obtained by spatially filtering the radio signals transmitted/received by an antenna array by assigning a weight to each antenna element according to the sensitivity pattern that we want to achieve. Owing to phase and gain differences between signals at different antennas, any given set of weights produces constructive interference in some directions and destructive interference in other directions. If the weights are constant and predefined, we refer to "fixed beamforming" or "switched-beam smart antennas"; otherwise, if the weights are selected adaptively on the basis of current measurements, we refer to "adaptive beamforming" or "adaptive-array smart antennas". In the latter case, location information can be used to provide adaptive coverage by identifying areas where there is a sudden need to provide higher capacity (e.g., in a stadium that is hosting the finals of the Olympic Games).

2.4.2.2 Power control

The problem that power control (PC) deals with is how to achieve the minimum transmission power level for the elements of the network while guaranteeing the desired QoS for each user, usually expressed in terms of a target signal-to-interference-plus-noise ratio (SINR) for each user terminal. In particular, when PC is applied to the downlink, the aim is to minimize multicell interference by balancing the transmission powers at the BSs or APs; when it is applied to the uplink, the aim is to minimize the in-band interference and the energy consumption by balancing the transmission powers at the user terminals. Usually, PC algorithms rely solely on signal strength measurements, either obtained at the BS or AP or reported back by the user terminal. The use of location awareness could greatly enhance their functioning. For example, as Chaudhari (2007) describes, "If a user is in a 'bad zone', the base station can keep its transmission power high even when the mobile is reporting a good signal strength. This will avoid 'spikes' in power level similar to the 'ping-pong' effect in handovers."

2.4.2.3 Packet scheduling

Packet scheduling is an RRM function that is controlled in downlink by the BS or AP and in uplink by the mobile station. The primary aim of a packet scheduler is to exploit the available physical resources (i.e., time, frequency, space and code) in such a way as to maximize the overall capacity while meeting the QoS requirements (usually defined in terms of a set of predefined parameters or priority levels) for the various links active for each user. Advanced packet-scheduling algorithms use important contingent information such as measures of instantaneous channel conditions (in terms of SINR) to allocate the available resources. For example, in

OFDMA systems (see Chapter 3), the packet scheduler at the BS or the AP may assign a certain subchannel to the terminal that is currently experiencing the best channel conditions, which can therefore use the highest modulation and coding scheme to transmit on that band. However, while on the one hand this approach maximizes the overall capacity of the system, on the other hand it results in an unfair distribution of resources among the users, as terminals closer to the BS or AP will be served more often than those at the cell edge. A proportional fair metric, which considers not only the current achievable rates but also the history of the rates achieved in the past by the users, has been therefore introduced in the literature to ensure a more democratic distribution of resources (Wengerter et al. 2005). However, in general, the aforementioned scheduling mechanisms can be categorized under the label of "reactive" mechanisms, since they are based on current information about the user. As outlined by Chaudhari (2007), "Sometimes the base station can allocate a significant chunk of bandwidth to a user at the expense of others only to find out that that terminal soon reports back bad channel conditions or, even worse, that the channel conditions of that terminal swing from one extreme to another, thereby driving the packet scheduler in a 'hunt mode', where it constantly changes its forward link transmission parameters." In contrast, "proactive" mechanisms, which feed back not only information about channel conditions but also about the current location and speed of the terminal, can be used in combination with database topological data to predict the future channel conditions and thereby perform a more effective scheduling choice.

2.4.2.4 Handover

Handover is the process that enables the mobility of a user within a network. Basically, it can be either intracell, when the user is moving from one sector to another in the same cell, or intercell, when the user is moving out of their home BS's coverage and has to be handed over to a neighboring BS. In the latter case, if both outgoing and incoming connections are maintained continuously, the handover is refer to as "soft" to differentiate from a "hard" handover, where there is a temporary disruption of the ongoing communication. Usually, the criteria on which the handover decision is based are (Chaudhari 2007) signal strength measurements (if these are provided by the user terminal, the handover is referred to as mobile-assisted), a list of neighbors, the traffic distribution, network capacity, bandwidth and some others. However, the signal strength might give unreliable indications, owing to its location-dependent variability. For example, if a user is on the borderline between several cells, the terminal will report back measurements that will often trigger a handover, thus creating the so-called "ping-pong" effect. In contrast, if those measurements are combined with location information, the aforementioned effect can be avoided by delaying the handover to a more appropriate moment. Further benefits deriving from such a combination are found in the following applications (Chaudhari 2007). (1) *Interfrequency handover*, when a user is moving to an area where a different carrier frequency has to be used. (2) *Vertical handover*, when a user is moving

to an area where there is a different type of network in use (e.g., from UMTS to WLAN). (3) *High-speed microhandovers*: if a user is moving at high speed to an area where there are many microcells (overlapping with the current macrocell), it is not convenient to continuously perform handovers but is better to keep with the same macro-BS. (4) *Handover prioritization*: if the core network has to handle many handover requests coming from many BSs in the network, direction and speed information can be used to prioritize the more time-critical handovers.

Finally, note that other RRM functions are involved when handover is performed, and thereby they can be location-aided:

- **Cell selection.** This is the operation that associates a user with a certain cell out of a set of possible candidates. Usually, the metric used to take this decision is based on signal strength measurements performed between the terminal and the BSs or APs. If cell selection is done with the aid of location information, situations in which certain users suffer from poor quality of service can be avoided. For example, if a user is moving towards an area covered by two APs with respect to which the user's terminal is experiencing slightly different channel conditions, an estimation of the user's direction and speed might indicate the cell with lower signal quality as the best choice for serving the user over a certain time period (Chaudhari 2007).

- **Admission control and load control.** These are processes that prevent a network from being overloaded, thus helping to avoid dropped calls, falling QoS, etc. In particular, while admission control aims to avoid congestion by regulating the total number of users who have permission to access the network, load control has to counteract overload situations that the admission control was not able to prevent (e.g., an overload due to a sudden increase in the interference level). If location-related information is available on the network side, these functions can be helped in their tasks by predicting future congestion. For example, speed and direction could indicate that a user is moving towards a crowded stadium, and therefore admission control and load control will kick in to accommodate the incoming user under the jurisdiction of a BS which is less congested, thus also taking part in performing a load-balancing action.

2.5 Conclusions

In this chapter, we have given an overview of the application areas of positioning, namely location-based services and location-based optimization. According to one of the definitions given in the chapter, an LBS is "an information service accessible with a mobile device through a mobile network, utilizing an ability to make use of the location of the mobile device itself" (Virrantaus et al. 2001). Intuitively, this opens up the possibility of advanced existing services and a previously unforeseen panorama of new appealing services (e.g., proximity matching, and gaming in outdoor scenarios). In particular, some of the "killer applications" envisioned in the chapter could result

from the enabling of indoor positioning, which the current technologies still struggle to provide. In contrast to location-based services, location-based optimization does not imply any active behavior on the user side; instead, it aims at enhancing network performance in terms of coverage, capacity, energy consumption, etc. Such enhancement is in many cases transparent to the user, who will perceive it in terms of higher QoS.

3

Basics of Wireless Communications for Positioning

Nicola Marchetti

Center for TeleInFrastruktur (CTIF), Aalborg University, Aalborg, Denmark

3.1 Introduction

In this chapter, the basics of wireless communications and their relation to mobile positioning are introduced. First, in Section 3.2, the radio propagation scenario is described by classifying the various physical phenomena that underlie wireless communications in terms of large-scale fading (i.e., path loss and shadowing) and small-scale fading (i.e., multipath propagation). Several positioning algorithms are then briefly reviewed. Then, in Section 3.3, the possibilities and challenges of multiple-antenna techniques are introduced, focusing in particular on the two big families of *spatial diversity* and *spatial multiplexing* and on their achievable gains, namely array gain, diversity gain, multiplexing gain and interference reduction. The other important family of multiantenna techniques, i.e., beamforming, is discussed later, when we treat space division multiple access (SDMA), an access technique that is based on beamforming. Some considerations of the possibilities offered by multiple-input–multiple-output (MIMO) techniques in terms of positioning are provided. Next, Section 3.4 overviews modulation and access techniques, focusing in particular on orthogonal frequency division multiplexing (OFDM) and orthogonal frequency division multiple access (OFDMA), which are the most relevant techniques for future wireless positioning systems. Another modulation

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

technique, i.e., spread spectrum, and access techniques such as time division multiple access (TDMA), code division multiple access (CDMA) and Carrier sense multiple-access (CSMA)/collision avoidance (CA) are also described. Then, Section 3.5 illustrates some radio resource management (RRM) techniques, namely handoff and interference management, channel reuse, channel reservation, and power control, and illustrates how localization can benefit from them and vice versa. In Sections 3.6 and 3.7, we outline two emerging areas within wireless communications, i.e., cooperative communications and cognitive radio technology, and their interconnections with localization. Section 3.8 concludes the chapter.

3.2 Radio Propagation

The received signal power varies as a function of space, frequency and time. The variation can be classified as either large-scale or small-scale fading.

Large-scale fading is dealt by a propagation model that predicts the mean received signal strength for a given transmitter–receiver separation. Such a model gives an average for measurements from 4λ to 40λ (Rappaport 1996), where λ is the wavelength. This is useful for estimating the coverage area. Large-scale fading can, in turn, be classified into path loss and shadowing. Path loss deals with the propagation loss due to the distance between transmitter and receiver, while shadowing describes variations in the average signal strength due to varying environmental clutter at different locations.

Small-scale fading deals with signal strength characteristics within a small distance from the receiver location. In such a region of space, the average signal strength remains constant. Multipath propagation of electromagnetic waves is the main cause of the effects considered here. Small-scale fading includes variations in time and space, and frequency-selective fading.

For each type of fading, there are two kinds of conditions, one when the variability is high and the other when the variability is low over the observation interval. Thus, the signal strength at a particular location depends on both the large-scale fading and the small-scale fading. As the receiver moves, the instantaneous power of the received signal varies rapidly, giving rise to small-scale fading. As the distance between the transmitter and the receiver increases, the local average of the received signal power decreases gradually and can be predicted by the large-scale-fading statistics. The phenomenon of combined large- and small-scale fading is represented in Figure 3.1.

Three main factors influence radio wave propagation: reflection, diffraction and scattering. *Reflection* happens when the electromagnetic waves impinge upon a surface that has dimensions much larger than their wavelength. *Diffraction* is due to the effect of sharp edges in the path of the radio waves between the transmitter and the receiver. *Scattering* is caused when the electromagnetic waves encounter objects of dimensions much smaller than their wavelength.

Most radio propagation models use a combination of empirical and analytical methods. The empirical approach is based on fitting curves or analytical expressions

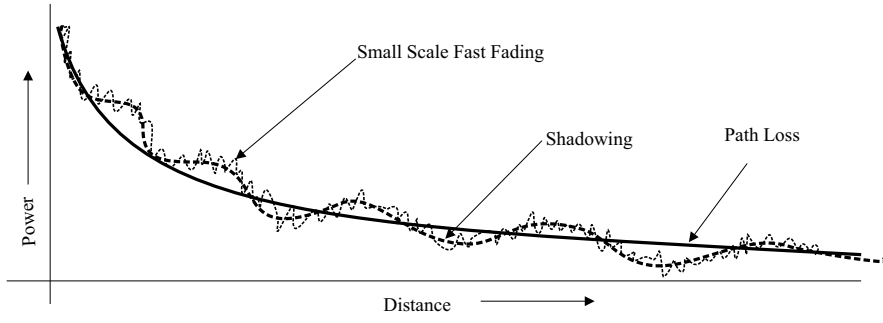


Figure 3.1 Propagation loss (Prasad et al. 2009).

that recreate a set of measured data. This has the advantage of implicitly taking into account all of the propagation factors. The validity of an empirical model for transmitter frequencies or environments other than those used to derive the model can be established only by use of additional measured data for the new environment and frequencies (Rappaport 1996).

3.2.1 Path loss

There are many models for predicting the path loss. such as the Hata–Okumura model and the COST231–Hata model (Rappaport 1996). Both of these models are for frequency ranges up to 2 GHz. To overcome this limitation, a model was proposed in Erceg et al. (1999a). For a close distance[1] $d_0 = 100$ m, the median path loss (Δp_{loss}) in dB is given by

$$\Delta p_{\text{loss}} = \alpha_{\text{loss}} + 10\beta \log\left(\frac{d}{d_0}\right) + \eta_{\text{shad}}, \quad d > d_0, \tag{3.1}$$

where η_{shad} is a random variable addressed in Section 3.2.2 and

$$\alpha_{\text{loss}} = 20 \log\left(\frac{4\pi d_0}{\lambda}\right), \tag{3.2}$$

η_{shad} is the shadowing factor, which follows a log-normal distribution with a typical value around 6 dB, and β is the path loss exponent; β is given by

$$\beta = a - bh_{\text{bs}} + c/h_{\text{bs}}, \quad 10 \text{ m} < h_{\text{bs}} < 80 \text{ m}, \tag{3.3}$$

where the values of a, b and c depend on the type of terrain (Erceg et al. 1999a) and h_{bs} is the BS antenna height in meters. The model in Equation (3.1) was proposed for

[1]The Friis free-space model, which is the basis for large-scale propagation models, is valid for values of d which are in the far field of the transmitting antenna. Therefore, large-scale propagation models (Rappaport 1996) use a "close distance" d_0, also known as the received power reference point. The path loss $\Delta p_{\text{loss}}(d)$ may be calculated at any distance $d \geq d_0$ in relation to the received power at d_0.

a receiver antenna height of 2 m and an operating frequency of 2 GHz. A correction factor for other frequencies and antenna heights can be found in Erceg (2001), and the modified path loss is then

$$\Delta p_{\text{mod}} = \Delta p_{\text{loss}} + \Delta p_{\text{freq}} + \Delta p_{\text{bsh}}, \tag{3.4}$$

where Δp_{loss} is the path loss given by Equation (3.1), Δp_{freq} is a frequency correction term given by $6\log(f/2000)$, where f is the frequency in MHz, and Δp_{bsh} is the receiver antenna height correction term given by $-10.8\log(h/2)$, where h is the new receiver antenna height, such that $2\text{ m} < h < 8\text{ m}$.

Considering only the path loss in Equation (3.1), it is possible to define the received power p based on Δp_{loss}, the transmit power p_{tx} and the transmitter and receiver antenna gains G_{tx} and G_{rx}, respectively:

$$p = \alpha - 10\beta\log(d) + \eta_{\text{shad}}, \tag{3.5}$$

where

$$\alpha \simeq p_{\text{tx}} + G_{\text{tx}} + G_{\text{rx}} - 20\log\left(\frac{4\pi}{\lambda}\right). \tag{3.6}$$

The propagation model used for long-term evolution (LTE) for 3GPP can be found in 3GPP (2006), where different parameters are used for different channel conditions and cell orientations.

3.2.2 Shadowing

The path loss model does not capture the varying environmental clutter at different locations. However, measurements have shown that for any value of d, the path loss Δp_{loss} at a particular location is random and log-normally distributed (i.e., normally distributed when measured in dB) around the mean path loss value (Rappaport 1996). Since the surrounding environmental clutter may be different at different locations, the path loss will be different from the average value predicted by Equation (3.4). This variation is mainly due to reflection and diffraction by interfering objects along the path of the traveling signal, and it is an additive term to the path loss but with random values. This phenomenon is called shadowing. Therefore, the received power loss expression results in

$$p = \alpha - 10\beta\log(d) + \eta_{\text{shad}}, \quad \eta_{\text{shad}} \sim \text{Norm}(0, \sigma_{\text{shad}}), \tag{3.7}$$

where η_{shad} is a zero-mean Gaussian-distributed random variable expressed in dB, with standard deviation σ_{shad}.

3.2.3 Small-scale fading

In small-scale fading, the signal varies rapidly over a short distance. This variation is caused by multipath propagation of the received signal and by Doppler frequency

shifts. The channel impulse response $h(t, \mathsf{t})$ is a function of two variables: the time t and the delay t (Proakis 1995).

Owing to reflecting objects such as buildings, hills and trees, delayed versions of the transmitted signal, each with a different amplitude ($\mathsf{a}^{(m)}$ in Equation (3.8) below) and phase ($\theta^{(m)}$ in Equation (3.8)), arrive at the receiver with different delays ($\mathsf{t}^{(m)}$ in Equation (3.8)). These parameters (amplitude, phase and delay) are random variables and can be characterized by a channel impulse response. If a unit impulse is transmitted and there are n_{se} scattering elements, then the receiver will receive n_{se} different signals. Therefore, the channel impulse response is the sum of these n_{se} scattered signals, as given below (Molisch 2005):

$$h(t, \mathsf{t}) = \sum_{m=1}^{n_{\mathrm{se}}} \mathsf{a}^{(m)} \delta(t - \mathsf{t}^{(m)}) \exp(-j\theta^{(m)}). \tag{3.8}$$

As is noticeable from Equation (3.8), the channel impulse response is a function of time, frequency and space (Lawton et al. 1991). In this equation, $\delta(\bullet)$ is the Dirac delta function and j is the imaginary unit of complex numbers.

3.2.3.1 Multipath fading

Multipath propagation gives rise to small-scale fading in the time and frequency domains. The multipath properties of a given environment are usually characterized by the power delay profile. The power delay profile is constructed from the values of the average power for each path. When one of the multipath components has a much higher power than the other components, then one is dealing with a Ricean channel. When this does not happen, usually in non-line-of-sight (NLOS) scenarios, one is dealing with a Rayleigh channel. One of the most commonly used models assumes that the power delay profile of a typical Rayleigh multipath channel has an exponential decay (Erceg et al. 1999b). There are several other models which consider the cluster effect; in these models, there is a double exponential decay, where each multipath is followed by a sequence of multipaths with a steeper decay constant during a very short interval. Another model for the delay profile has the first few taps with same average power followed by an exponential decay (Witrisal 2002).

3.2.4 Radio propagation and mobile positioning

Several mobile positioning methods, ranging from low-accuracy methods based on cell identification to high-accuracy methods combining wireless network information and satellite positioning, have been proposed. All of these methods exploit measurements or estimated parameters which relate the MS position to the position of fixed reference points (FRPs), for example the positions of BSs, or to the specific behavior of the MS and its surrounding environment. For instance, while measurements of the angle of arrival (AOA) (Section 4.3.5) give the direction between the MS and the FRP, measurements of the time of arrival (TOA), time

difference of arrival (TDOA) and received signal strength (RSS) (Sections 4.3.2–4.3.4) provide information about the distance between the MS and the FRP (Miao 2007).

Formally, position estimation can be interpreted as a problem of solving nonlinear systems of equations. In general, the solutions are obtained by applying a stochastic gradient algorithm or by numerically approximating the nonlinear least-squares (NLLS) problem using Monte Carlo-based techniques. We shall now present an overview of the measurements used for positioning in most wireless networks, and positioning algorithms associated with the various measurements.

3.2.4.1 Measurements

Let $X = [x, y]^{\mathrm{T}}$ denote the 2D position of the mobile device. The ith known BS position is defined by $X_i = [x_i, y_i]^{\mathrm{T}}$. The BSs can move with time; this happens, for example in some ad hoc networks and sensor networks. A generic measurement z_i obtained at the ith BS is a function $h_i(X)$ of both the MS position and the ith BS position and is affected by noise η_i, i.e.,

$$z_i = h_i(X) + \eta_i. \tag{3.9}$$

As stated later in Section 4.3, typical measurements include RSS, TOA, TDOA and AOA measurements, database correlation (digital map information), and position estimates (Miao 2007).

3.2.4.2 Position estimation

Given a measurement vector $Z = [z_1, \ldots, z_n]^{\mathrm{T}}$ from n different BSs and the associated measurement transform $\mathcal{H}(X) = [h_1(X), \ldots, h_n(X)]^{\mathrm{T}}$ and assuming that the measurement noise vector is $\eta = [\eta_1, \ldots, \eta_n]^{\mathrm{T}} \sim \mathrm{Norm}(0, R)$, where $\mathrm{Norm}(0, R)$ denotes a normal distribution with zero mean and autocorrelation matrix R, a position estimation can be obtained by various algorithms, as described below. See Chapters 4 and 5 for further details.

3.2.4.2.1 Nonlinear least-squares algorithm When the measurement noise vector η is white, the NLLS position estimator provides an optimal approach equivalent to the maximum-likelihood (ML) method. The NLLS-based position estimate is obtained from

$$\hat{X}_{\mathrm{nlls}} = \arg\min_{X}\{(Z - \mathcal{H}(X))^{\mathrm{T}}(Z - \mathcal{H}(X))\}. \tag{3.10}$$

3.2.4.2.2 Weighted NLLS When η is a colored noise vector, the NLLS estimation can be improved by use of the weighted nonlinear least squares (WNLLS) estimate, formulated as

$$\hat{X}_{\mathrm{wnlls}} = \arg\min_{X}\{(Z - \mathcal{H}(X))^{\mathrm{T}} R^{-1}(Z - \mathcal{H}(X))\}. \tag{3.11}$$

3.2.4.2.3 ML estimation In the more general case where the Gaussian noise vector is dependent on the position, i.e., $\eta \sim \text{Norm}(0, R(X))$, the ML estimate is formulated as

$$\hat{X}_{\text{ml}} = \arg\min_{X}\{(Z - \mathcal{H}(X))^{\text{T}} R^{-1}(Z - \mathcal{H}(X)) + \log|R(X)|\}, \tag{3.12}$$

where the term $\log|R(X)|$ prevents the selection of positions with large uncertainty (i.e., large $\log|R(X)|$), which could be the case with the WNLLS algorithm.

In general, as there is no closed-form solution for the algorithms above, numerical search methods are natural choices, for example the steepest-gradient method and Newton-type methods (Dennis and Schnabel 1983; Peressini et al. 1988). These local search algorithms require a good initialization, otherwise there is a risk of reaching only a local optimum of the criterion function.

3.2.4.3 NLOS positioning error mitigation

All of the algorithms considered above assume that line-of-sight (LOS) propagation paths exist between the MS and the BSs. In the presence of NLOS propagation, the major positioning errors result from the measurement noise and the NLOS propagation error, which is the dominant factor (Caffery and Stüber 1998). Compared with an LOS situation, a time estimate in an NLOS situation will have a positive bias μ and probably a larger variance. The probability density function (PDF) of the NLOS measurement error can be described by a Gaussian distribution (Cong and Zhuang 2005) as

$$\eta_i \sim \text{Norm}(\mu, \sigma^2_{\text{nlos}}), \tag{3.13}$$

where μ defines the bias introduced by the NLOS propagation and σ^2_{nlos} denotes the variance of the measurement error. The PDF of the measurement error can be modeled by a mixture of two Gaussian distributions (Cong and Zhuang 2005) as

$$\eta_i \sim \gamma\ \text{Norm}(0, \sigma^2_{\text{los}}) + (1 - \gamma)\ \text{Norm}(\mu, \sigma^2_{\text{nlos}}), \tag{3.14}$$

where γ stands for the probability of the measurement η_i resulting from LOS paths. To mitigate the effects of NLOS paths, two methods are used: one is to use a robust error distribution as in Equation (3.14) to design the positioning algorithm, and the other is to include some searching mechanism to exclude the NLOS outliers (Chen 1999; Cong and Zhuang 2001; Xiong 1998). See Chapter 7 for further mitigation algorithms.

3.2.5 RSS-based positioning

In RSS-based positioning, the target MS measures the received signal strength from the surrounding BSs and the measured data is used to estimate the location of the MS (Soork et al. 2008). Let $X = [x, y]^{\text{T}}$ denote the target MS's coordinates and let $\{X_i = [x_i, y_i]^{\text{T}}, i = 1, \ldots, n\}$ denote the coordinates of the n BSs (i.e., $\text{BS}_1, \ldots, \text{BS}_n$).

The received signal strength from the ith BS (BS_i) is modeled, based on Equation (3.5), as

$$p_i = \alpha_i - 10\beta \log(d_i) + \eta_i^{\text{shad}}, \tag{3.15}$$

where d_i is the separation distance between BS_i and the MS, and p_i is the received power strength at BS_i. In order to estimate the position of the target MS using Equation (3.15), a least-squares objective function can be defined as

$$\mathcal{J}(X) = \sum_{i=1}^{n} (\eta_i^{\text{shad}})^2 = \sum_{i=1}^{n} [p_i - \alpha_i + 10\beta \log(d_i)]^2, \tag{3.16}$$

or, in matrix form, as

$$\mathcal{J}(X) = [Z - \mathcal{H}(X)]^{\text{T}} [Z - \mathcal{H}(X)], \tag{3.17}$$

where

$$Z = [p_1, \ldots, p_n]^{\text{T}}, \tag{3.18}$$

$$\mathcal{H}(X) = [\alpha_1 - 10\beta \log(d_1), \ldots, \alpha_n - 10\beta \log(d_n)]^{\text{T}}. \tag{3.19}$$

The position estimate will be the optimum point of the nonlinear function $\mathcal{J}(X)$ and is calculated numerically. Numerical algorithms start from an initial point and iteratively update it to reach the solution. In order to avoid convergence of the numerical algorithm to a local minimum, the starting point must be chosen not far away from the optimum. In order to specify a suitable initial point, one first calculates an estimate of the distance between the MS and all BSs. For further information about least-squares methods and the associated formalism, see Section 5.2.

3.3 Multiple-antenna Techniques

The use of multiple antennas is one of the key tools for mitigating the negative effects of the wireless channel, thereby providing better link quality and/or a higher data rate without consuming extra bandwidth or transmitting power. The use of multiple antennas at the transmitter and/or receiver provides various benefits, i.e., array gain, interference reduction, diversity gain and/or multiplexing gain (Paulraj et al. 2003). The combination of multiantenna techniques with OFDM can be very beneficial, since OFDM has the capability of turning a frequency-selective MIMO fading channel into multiple flat fading channels (Bolcskei et al. 2002). This renders the multichannel equalization particularly simple, since for each OFDM subcarrier only a constant matrix needs to be inverted.

Multiantenna techniques can be broadly classified into three categories: (1) smart array processing, such as beamforming (BF), aimed at increasing received power and rejecting unwanted interference; (2) spatial diversity (SD), employed to mitigate fading and enhance link reliability; and (3) spatial multiplexing, used to boost the data rate.

3.3.1 Spatial diversity

Spatial diversity is a method in which two or more antennas physically separated from each other are used in order to obtain independent versions of the received signal. These antennas can be located at the receiver (*receive diversity*), at the transmitter (*transmit diversity*) or at both.

The level of cross-correlation between the signals transmitted or received by antennas belonging to the same device depends on the distance between the antennas. In general, the minimum distance between the antennas must be bigger than the coherence distance of the channel, so that the spatial fadings are uncorrelated (Paulraj et al. 2003). Ideally, an antenna separation of half of the wavelength ($\lambda/2$) should be sufficient (Jakes 1994). If the antenna separation is sufficient, spatial diversity can be employed to reduce fading. Spatial diversity is relatively simple to implement and does not require additional frequency spectrum; the downside of spatial diversity, as for all multiantenna techniques, is the ensemble of cost, size, power consumption and complexity, due to the fact of having additional antennas.

When multiple copies of the transmitted signal arrive at the receiver, they have to be precombined in order to obtain a diversity gain.

3.3.2 Spatial multiplexing

In contrast to diversity techniques, spatial multiplexing is aimed at increasing the data rate of the system. Spatial multiplexing techniques require multiple antennas at both the transmitter and the receiver; such a system is often referred to as a MIMO system, and is shown in Figure 3.2. The numbers of transmit and receive antennas are denoted by n_{tx} and n_{rx}, respectively. At the transmitter, n_{tx} independent data streams, $\{s_0^{\text{tx}}(t), s_1^{\text{tx}}(t), \ldots, s_{n_{\text{tx}}-1}^{\text{tx}}(t)\}$, are sent from different antennas at the same time and at the same frequency. For the sake of simplicity, flat fading channels are assumed between pairs of transmit and receive antennas here. At the qth receive antenna, the independent signals $\{s_0^{\text{tx}}(t), s_1^{\text{tx}}(t), \ldots, s_{n_{\text{tx}}-1}^{\text{tx}}(t)\}$ are linearly combined to produce the received signal $s_q^{\text{rx}}(t)$, which is given by

$$s_q^{\text{rx}}(t) = h_{q,0}s_0^{\text{tx}}(t) + h_{q,1}s_1^{\text{tx}}(t) + \cdots + h_{q,n_{\text{tx}}-1}s_{n_{\text{tx}}-1}^{\text{tx}}(t) + \eta_q^{\text{therm}}(t), \tag{3.20}$$

for $q = 0, 1, \ldots, n_{\text{rx}} - 1$. In Equation (3.20), $h_{q,p}$ denotes the envelope of the flat fading channel between the pth transmit and the qth receive antenna, and $\eta_q^{\text{therm}}(t)$ is the thermal noise at the qth receive antenna. In a rich scattering environment, the channels between the pairs of transmit and receive antennas experience different multipath conditions, and thus are independent with respect to each other. The independence of the channels implies that the receiver has to estimate $\{s_0^{\text{tx}}(t), s_1^{\text{tx}}(t), \ldots, s_{n_{\text{tx}}-1}^{\text{tx}}(t)\}$ from $\{s_0^{\text{rx}}(t), s_1^{\text{rx}}(t), \ldots, s_{n_{\text{rx}}-1}^{\text{rx}}(t)\}$ independently, provided that the channel responses are known (Paulraj et al. 2003). The problem is equivalent to solving a set of n_{rx} linear equations in n_{tx} unknown variables, which can be solved if $n_{\text{rx}} \geq n_{\text{tx}}$.

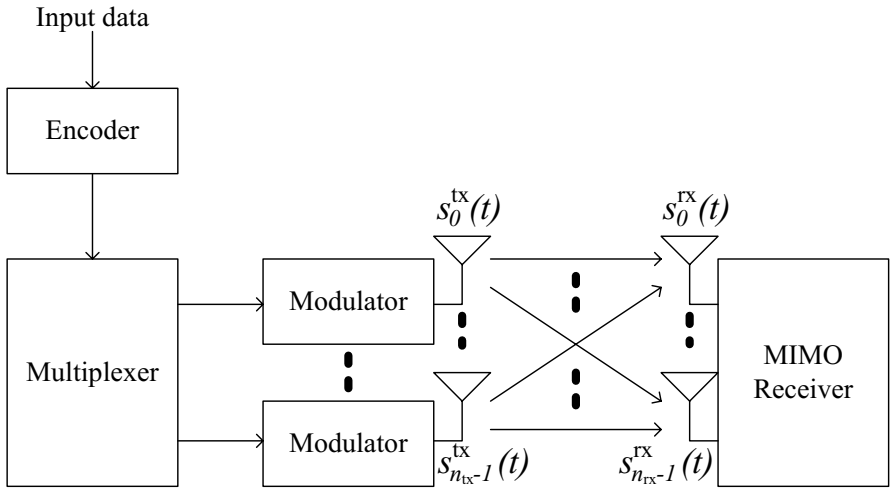


Figure 3.2 A spatial multiplexing system.

As a result, spatial multiplexing promises higher throughput compared with single-input–single-output (SISO) systems, without the need to increase the system bandwidth or the transmit power. As Foschini and Gans (1998) have shown, under certain circumstances the system capacity increases linearly with the number of antennas used.

3.3.3 Gains obtained by exploiting the spatial domain

Let us first examine the possible transceiver architecture configurations. SISO refers to a system with one transmit and one receive antenna; single-input–multiple-output (SIMO) refers to one antenna at the transmitter and multiple antennas at the receiver, and multiple-input–single-output (MISO) refers to the converse; and MIMO indicates an architecture with multiple antennas both at the transmitter and at the receiver.

Let us now discuss the possible gains that are achievable through the use of multiple antennas in a wireless system. In the following, we shall refer to the numbers of transmit and receive antennas as n_{tx} and n_{rx}, respectively.

3.3.3.1 Array gain

This term indicates the average increment in signal-to-noise ratio (SNR) derived from the effect of coherent combination of multiple antennas at the transmitter, receiver or both. In the case of SIMO, the signals arriving at the receive array have different amplitudes and phases: in this case, the receiver can combine the incoming signals coherently, such that the resulting signal is enhanced. This results in an average increase in signal power at the receiver, proportional to n_{rx}. In the case of systems

with multiple antennas at the transmitter, one requires knowledge about the channels at the transmitter to exploit array gain (Paulraj et al. 2003).

3.3.3.2 Diversity gain

In a wireless channel, the signal power is subject to fluctuations (this is also known as the fading phenomenon); when the signal power decreases by a considerable amount, the channel is said to be in a fade. The techniques designed to combat the fading problem are called diversity techniques.

When one is dealing with a SIMO channel, receive diversity techniques can be employed (Jakes 1994). The algorithms for these techniques exploit the fact that different receive antennas see independently faded versions of the same signal, and the receiver then combines these signals in such a way that the resultant signal has a reduced amplitude variation compared with the signals at each receive antenna. The diversity gain, also called the diversity order, is equal to the number of receive antennas in the SIMO case.

In the case of a MISO channel, we can apply transmit diversity techniques and obtain diversity gain if knowledge about the channel is available at the transmitter. Probably the best-known examples of the designs that have been proposed in the literature to obtain transmit diversity are space–time coding (STC) schemes (e.g., Alamouti 1998; Tarokh et al. 1998), which encode across the spatial domain (with the transmit antennas) and the time domain, and do not make use of any channel knowledge at the transmitter. If the links between all n_{tx} transmit antennas and the receive antenna experience independent fading conditions, then the diversity order of the MISO channel is equal to n_{tx}.

If one is dealing with a MIMO channel, one can combine the above-mentioned transmit and receive diversity schemes. If the links between each transmit–receive antenna pair fade independently, then the diversity order becomes $n_{\mathrm{tx}}n_{\mathrm{rx}}$ (the product of the numbers of transmit and receive antennas) (Paulraj et al. 2003).

3.3.3.3 Multiplexing gain

Spatial multiplexing offers a linear increment $\min(n_{\mathrm{tx}}, n_{\mathrm{rx}})$ of the capacity, with no extra bandwidth or power expense; this gain is called the multiplexing gain. Only by having multiple antennas both at the transmitter and at the receiver, i.e., by using a MIMO channel, is spatial multiplexing possible (Foschini 1996; Paulraj and Kailath 1994; Telatar 1999).

In the case of $n_{\mathrm{tx}} = n_{\mathrm{rx}} = 2$, i.e., a 2×2 MIMO channel, the bit stream to be transmitted is demultiplexed into two half-rate substreams, simultaneously modulated and transmitted from each transmit antenna. Under certain favorable channel conditions, the spatial signatures of these signals induced at the two receive antennas are well separated, and thus receiver having the knowledge about the channel can differentiate between the two co-channel signals and extract both of

them. At this point, demodulation yields the original substreams, which can be combined, giving back the original bit stream (Paulraj et al. 2003).

Transmit and receive diversity are means to combat fading. A different line of thought suggests that in a MIMO channel, fading can in fact be beneficial, leading to an increase in the degrees of freedom available for communication (Foschini 1996; Telatar 1999). Essentially, if the path gains between individual transmit–receive antenna pairs fade independently, multiple parallel spatial channels are created; as a consequence, in a rich scattering environment, the independent spatial channels can be exploited to send multiple signals in parallel, at the same time and frequency, resulting in higher spectral efficiency, which in turn means that the data rate can be increased, given the same bandwidth. This effect is called spatial multiplexing (Heath and Paulraj 2000) and is particularly useful in the high-SNR regime, where the system is degrees-of-freedom-limited rather than power-limited. Foschini (1996) has shown that in the high-SNR regime, the capacity of a channel with n_{tx} transmit antennas, n_{rx} receive antennas and i.i.d. Rayleigh-faded gains between each antenna pair is given by

$$\mathcal{C}(\rho) = \min(n_{\text{tx}}, n_{\text{rx}}) \log \rho + O(1), \tag{3.21}$$

where ρ is the SNR and $O(1)$ means that at infinity this component behaves like a constant. The number of degrees of freedom is thus the minimum of n_{tx} and n_{rx}.

3.3.3.4 Interference reduction

Frequency reuse in wireless channels gives rise to co-channel interference; a possible way to limit this problem is to use multiple antennas to distinguish the spatial signature of the desired signal from the signatures of the interfering co-channel signals. Some knowledge about the channel of the desired signal is required to perform interference reduction.

Also, one can implement interference reduction or avoidance techniques at the transmitter; the objective in this case is to maximize the energy sent to the desired user and, at the same time, to minimize the energy sent to the co-channel users. Some positive effects of interference reduction are the possibility of using aggressive reuse factors and an improvement of network capacity (Paulraj et al. 2003).

3.3.4 MIMO and mobile positioning

During recent years various location technologies have been invented, using either cellular network-based, mobile-based or hybrid approaches. The most known and widely used positioning technique is the GPS technique, based on measurements of TOA. The propagation times of signals from satellites at known locations are measured simultaneously, and the distance between a satellite and a user receiver is obtained by multiplying the propagation time by the speed of light, assuming LOS conditions. In most applications, however, the LOS signal is followed by multipath components that arrive at the receiver with a delay, introducing significant errors into the LOS-path time of arrival and into the gain estimation, especially in urban

environments, which are characterized by many reflections from buildings and other objects. MIMO systems can use information from multipath components to improve the accuracy of the estimation. The key issue for positioning in MIMO settings is selecting an appropriate model for the propagation channel. With the proper channel model, the location problem can be solved using sequential Monte Carlo methods, also known as particle filtering (Doucet et al. 2001).

3.4 Modulation and Multiple-access Techniques

3.4.1 Modulation techniques

3.4.1.1 OFDM

The nature of future wireless applications will lead to an increasing demand for high data rates. OFDM, a special form of multicarrier transmission where all of the subcarriers are orthogonal to each other, promises high-user-data-rate transmission capabilities at reasonable complexity and precision.

At high data rates, the distortion of the data caused by the channel is very significant, and it is almost impossible to recover the transmitted data with a simple receiver. A very complex receiver structure, which makes use of computationally expensive equalization and channel estimation algorithms to correctly estimate the channel, is needed in order to use estimations in combination with the received data and recover the originally transmitted data. OFDM can drastically simplify the equalization problem by turning a frequency-selective channel into multiple flat channels; only a simple one-tap equalizer is needed in this case to estimate the channel and thus recover the data.

Future telecommunications systems must be spectrally efficient to support a large number of users requiring high-data-rate multimedia communications; in this context, OFDM proves to be useful, as it uses the available spectrum very efficiently. Therefore, pure OFDM (or hybrid OFDM-based modulation schemes) seems at the moment to be the best candidate underlying technology for the multiple-access techniques of the future generation of telecommunications systems.

The transceiver chain for an OFDM system is shown in Figure 3.3. In an OFDM transmitter, forward-error-correction (FEC) coding and interleaving are added to obtain robustness against burst errors. An OFDM system with this addition is referred to as a system using coded orthogonal frequency division multiplexing (COFDM). In the digital domain, binary input data is collected and FEC-coded with schemes such as convolutional codes. The coded bit stream is interleaved to obtain diversity gain. After this, a group of channel-coded bits is gathered together (one bit per symbol for binary phase shift keying (BPSK) modulation, two for quadrature phase shift keying (QPSK), four for 16-quadrature amplitude modulation (QAM), etc.) and mapped to the corresponding constellation points. Then, the data is represented as complex numbers, which are streamed serially. Known pilot symbols mapped with known modulation schemes can be inserted at this point. A serial-to-parallel converter

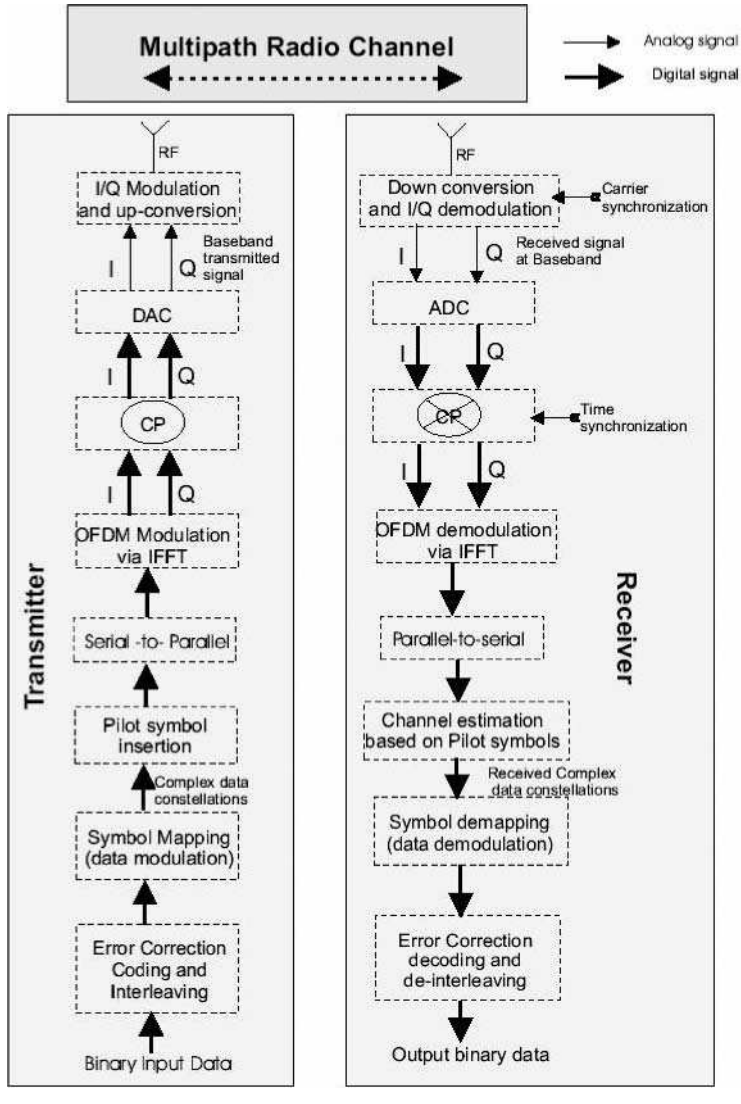


Figure 3.3 OFDM transceiver chain (Prasad et al. 2009).

is then applied, and an inverse fast Fourier transform (IFFT) operation is performed on the parallel complex data. The transformed data is grouped together again, in accordance with the number of transmission subcarriers required. A cyclic prefix is inserted into every block of data according to the system specification and the data is demultiplexed in a serial fashion. At this point, the data is OFDM-modulated and ready to be transmitted. A digital-to-analog converter (DAC) is used to transform the time-domain digital data to analog data. Radio frequency (RF) modulation is then performed, and the signal is up-converted to the transmission frequency.

After the transmission of the OFDM signal from the transmitter antenna, the signals go through the ever-unpredictable wireless channel. After receiving the signal, the receiver down-converts it to an intermediate frequency and then converts it to the digital domain by using an analog-to-digital converter (ADC). At the time of down-conversion of the received signal, carrier frequency synchronization is performed, and after ADC conversion, symbol-timing synchronization is carried out. A fast Fourier transform (FFT) block is used to demodulate the OFDM signal. After that, channel estimation is performed using the demodulated pilots. Using the resulting estimations, the complex received data is obtained and then demapped according to the transmission constellation diagram used. At this point, FEC de-interleaving and decoding take place, aiming at recovering the originally transmitted bit stream.

3.4.1.2 Spread spectrum

Spread spectrum is a transmission technique in which a pseudo-noise code, independent of the information data, is used as a modulation waveform to spread the signal energy over a bandwidth much greater than the signal information bandwidth. At the receiver, the signal is despread using a synchronized replica of the pseudo-noise code (Meel 1999). Hence, spreading does not economize on the limited frequency resource. However, this overuse is well compensated by the possibility that it provides for many users to share the enlarged frequency band.

There are many benefits associated with the spread-spectrum technology, the most important of which is resistance to interference. Interference and jamming signals are rejected because they do not contain the spread-spectrum key: only the desired signal, which has such a key, will be seen at the receiver when the despreading operation is done. This means that one can practically ignore interference, whether narrowband or wideband, if it does not include the key used in the despreading operation. This rejection also applies to other spread-spectrum signals that do not have the right key. This leads to the possibility of having different spread-spectrum communications active simultaneously in the same band, such as in the case of CDMA.

Resistance to interception is the second advantage provided by spread-spectrum techniques; since unauthorized listeners do not have the key used to spread the original signal, they cannot decode the signal. Without the right key, the spread-spectrum signal appears as noise or interference.

Spread spectrum also provides resistance to multipath fading. Since the despreading process is synchronized to the direct-path signal, the reflected-path signal is rejected even though it contains the same key.

The main characteristic of spread spectrum is the presence of a key or code, which must be known in advance by the transmitter and receiver(s). In modern communications, these codes are digital sequences that must be as long and as random as possible to appear “noise-like”; however, such codes must remain reproducible, otherwise the receiver will not be able to extract the message that has been sent. Such a nearly random code is called a pseudo-random number (PRN) or sequence, and is usually generated by using a feedback shift register.

Different spread-spectrum techniques are distinguished according to the point in the system at which a PRN is inserted into the communication channel. If the PRN is inserted at the data level, the technique is called direct-sequence spread spectrum (DSSS). If the PRN acts at the carrier-frequency level, the technique is called frequency-hopping spread spectrum (FHSS). Finally, if the PRN is applied at the local-oscillator stage, FHSS PRN codes force the carrier to change (hop) according to the pseudo-random sequence. Analogously, if the PRN acts as an on/off gate to the transmitted signal, the technique is referred to as time-hopping spread spectrum (THSS). One can also mix the above techniques, forming a hybrid spread-spectrum technique. DSSS and FHSS are the techniques most in use today (Maxim 2003).

3.4.2 Multiple-access techniques

3.4.2.1 TDMA

In TDMA, the same bandwidth is shared by all users, but each of them transmits only in certain time slots, thus sharing the channel between all users over time (hence the words "time division") (Celusion 2009); this multiple-access technique is used in GSM.

3.4.2.2 FDMA/OFDMA

In frequency division multiple access (FDMA), the available bandwidth is divided into channels, each being allocated to a different user for transmission (Celusion 2009). As an example of FDMA, we shall mention OFDMA, which is used in the Worldwide Interoperability for Microwave Access (WiMAX) system and LTE.

In OFDMA, the available subcarriers are distributed among all of the users for transmission. The subcarrier assignment is made for the lifetime of the user connection or at least for a considerable time frame. This scheme was first proposed for cable television systems (Sari and Karam 1996) and later on adopted for wireless communication systems. OFDMA can support a number of identical or different user data rates, for example by assigning a different number of subcarriers to each user. Based on the conditions of the subcarriers or subchannels (groups of subcarriers), several different baseband modulation schemes can be used for the individual subcarriers/subchannels, for example QPSK, 16-QAM or 64-QAM. This is referred to as adaptive subcarrier/subchannel, bit, and power allocation or QoS-based allocation (Kivanc and Liu 2000; Rhee and Cioffi 2000; Wong et al. 1999a,b). In OFDMA, frequency hopping can be employed to provide security and resilience to intercell interference.

In OFDMA, the granularity of resource allocation is higher than that in OFDM-TDMA, where flexibility can be accomplished by suitably choosing the subcarriers associated with each user at different times. Here, the fact that each user experiences a different radio channel can be exploited by allocating only "good" subcarriers with high SNR to each user. In addition to this, the number of subchannels for a specific

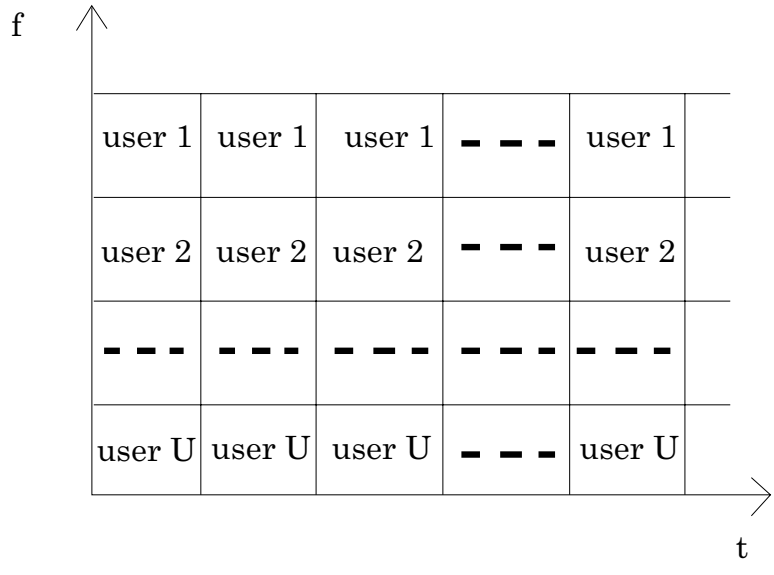


Figure 3.4 An example of a time–frequency user allocation diagram for the OFDMA scheme.

user can be varied according to the required data rate. Thus, a multirate type of system can be achieved without increasing the complexity of the system too much.

In principle, subcarriers can be assigned arbitrarily among the subchannels. An example of an assignment of subcarriers to users is illustrated in Figure 3.4.

3.4.2.3 CDMA

In a CDMA system, as is utilized by third generation (3G) systems, all users occupy the same frequency and are separated from each other by means of a code. Each user is assigned a code, which is used to transform the user's signal into a spread-spectrum-coded version of the user's data stream. The receiver uses the same spreading code in order to transform the spread-spectrum signal back into the original user's data stream (Celusion 2009).

3.4.2.4 SDMA

SDMA is a multiple-antenna-based scheme which enables access by identifying the user location; this enhances the efficiency of network bandwidth utilization. SDMA can be applied on top of most of the well-known mobile access schemes such as CDMA, TDMA and FDMA.

SDMA enables the channeling of radio signals based on the mobile devices' locations. In this way, SDMA not only protects the quality of radio signals, avoiding

interference from adjacent cells, but also saves unnecessary signal transmission in areas where mobile devices are not currently active or available.

SDMA works through simultaneous transmission of radiation streams at the same frequency from a transmitter that is equipped with multiple antennas (potentially, SDMA can be enhanced if the receiver is also equipped with multiple antennas) (SDMA Toolbox for IT 2009).

3.4.2.5 CSMA/CA

Carrier sense multiple-access (CSMA) protocols are well known in the industry, the most popular being Ethernet, which is a CSMA/collision detection (CD) protocol. A CSMA protocol works as follows: a station desiring to transmit senses the medium; if the medium is not busy (i.e., if some other station is not transmitting), the station is allowed to transmit. This multiple-access technique is utilized by Wireless Fidelity (Wi-Fi).

These kinds of protocols are very effective when the medium is not heavily loaded, since they allow transmission with minimum delay. However, there is always a chance of collisions, which are caused when other stations have sensed the medium as being free and decided to transmit at the same time. In the case of Ethernet, these collisions are recognized by the transmitting stations, which perform a retransmission phase based on an exponential random backoff algorithm; for example, retransmissions are done after waiting after a failed transmission for a time that is chosen randomly from a set of possible waiting times, and the cardinality of this set increases exponentially with time. While these collision detection mechanisms work well in a wired local area network (LAN), they cannot be used in a wireless LAN (WLAN) environment, because of the following reasons:

- Implementing a collision detection mechanism would require the implementation of a full duplex radio, which would increase the price significantly.
- In a wireless environment, we cannot assume that all stations hear each other (which is the basic assumption of the collision detection scheme), and the fact that a station willing to transmit senses the medium as being free does not necessarily mean that the medium is actually free around the receiver area.

In order to overcome these problems, the 802.11 standard uses a CSMA/collision avoidance (CA) mechanism together with a positive acknowledge scheme, as follows. A station willing to transmit senses the medium; if the medium is busy, then the station defers its transmission. If the medium is free for a specified time, then the station is allowed to transmit; the receiving station will check the cyclic redundancy check (CRC) of the received packet and send an acknowledgment packet (ACK). The receipt of the acknowledgment will indicate to the transmitter that no collision has occurred. If the sender does not receive an acknowledgment, then it will retransmit the fragment until it is acknowledged, or discard the packet after a certain number of retransmissions.

In order to reduce the probability of two stations colliding owing to the fact that they cannot hear each other, a *virtual carrier sense* mechanism has been defined: a station willing to transmit a packet first transmits a short control packet called request to send (RTS), and the destination station responds (if the medium is free) with a response control packet called clear to send (CTS). All stations receiving either the RTS or the CTS set their virtual-carrier-sense indicator – called the network allocation vector (NAV) – for a given duration, and use this information together with the physical carrier sense when sensing the medium. This mechanism reduces the probability of a collision in the receiver area caused by a station that is "hidden" from the transmitter for the duration of the RTS transmission, as that station will hear the CTS and "reserve" the medium until the end of the transaction. It should also be noted that since the RTS and CTS are short frames, such a mechanism reduces the overhead of collisions, as the latter are recognized faster than they would be if the whole packet were to be transmitted (Brenner 1997).

3.4.3 OFDMA and mobile positioning

As an example of the relation of multiple-access techniques to mobile positioning, we shall consider an OFDMA-based system, namely WiMAX. In Jiao et al. (2008), a mobile WiMAX positioning scheme that adopts a layered infrastructure was proposed. There are three function modules, located in the physical layer (PHY), the medium access control layer (MAC) and the application layer, which communicate with each other through a cross-layer mechanism. The PHY module is the key part of the structure, which performs preamble detection and measurement in the MS and selects several approaches for positioning, for example TDOA. The MAC module performs signaling and provides opportunities for channel measurements according to positioning requests from the application layer. Position and optional velocity computations are performed in the application layer based on the measurements. Positioning algorithms are embedded either in the location computation server in the cognitive-service network or in the MS. The communication among the components in the application layer (including the communication between the MS or the cognitive-service network and the location computation server) follows the Mobile Positioning Protocol, which is an Internet-based protocol used by location-dependent applications to interface with a mobile positioning server. Through this protocol, it is possible to request the position of mobile terminals (Rios 2009).

3.5 Radio Resource Management and Mobile Positioning

3.5.1 Handoff, channel reuse and interference adaptation

One possible way to increase the capacity of a cellular system is to adapt to the dynamic changes in the traffic loads in the various cells and to adapt to real-time interference measurements. The aim of both of these types of adaptation is to enhance cell reuse without producing a detrimental level of co-channel interference

(Chiu and Bassiouni 2000). The average signal strength p (in linear units, i.e., watts) at a distance d from the transmitter is usually modeled as (Zander and Erikkson 1993)

$$p = \frac{10^{\alpha/10}}{d^{\beta}}, \tag{3.22}$$

where α and the path loss exponent β are taken from Equation (3.5). Assuming no noise this relation can be used to compute the carrier-to-interference ratio (CIR), written as C/I, for the downlink channel as follows. Let BS_0 represent the BS that currently communicates with the MS and let BS_i represent an interfering BS, i.e., a nearby BS which shares the use of the same downlink channel with BS_0. Then, the downlink C/I, without considering the effect of shadowing, is given by

$$\frac{C}{I} = \frac{d_0^{-\beta}}{\sum_{i \neq 0} d_i^{-\beta}} \geq \xi, \tag{3.23}$$

where d_i is the distance between BS_i and the MS, and ξ is a threshold. The above constraint implies that to achieve a good downlink connection, BS_0 should provide the highest average signal level for the MS. Moreover, the signal from BS_0 should be sufficiently strong with respect to the sum of the interference signals from the other BSs. A similar constraint applies to the uplink case.

3.5.1.1 Handoff prioritization

The mechanism of handoff in cellular networks transfers an ongoing call from the current cell to the next cell as the MS moves through the coverage area of the cellular system. A successful handoff provides continuation of the call, which is crucial to keeping a satisfactory perceived QoS (Chiu and Bassiouni 2000).

The notion of bandwidth reservation using static guard channels is well known (Hong and Rappaport 1986; Tripathi et al. 1998). The shadow cluster concept was proposed by Levine et al. (1997) to estimate future resource requirements and perform admission control, with the aim of limiting the handoff-dropping probability. The potential influence of the active MSs on the channel resources of each cell in the cluster is determined probabilistically by the BS of the current cell based on previous knowledge of the mobility pattern of the active MSs. Lu and Bharghavan (1996) explored mobility estimation for an indoor wireless system based on both mobile-specific and cell-specific observation histories. Talukdar et al. (1997) proposed a bandwidth reservation scheme that assumes that the mobility of an MS can be characterized by the set of cells that the MS is expected to visit during its connection's lifetime, reserving bandwidth in each of these cells, which is often excessive. Choi and Shin (1998) introduced a predictive and adaptive scheme which also performs bandwidth reservation for handoffs. The aggregate history of mobility observed at the cell level is used to predict probabilistically the direction of the MS and the time of the expected handoff.

One problem with all history-based schemes is the overhead related to the creation and update of traffic histories for the various cells. These histories are never fully reliable, since they continuously experience short-term changes (e.g., diversion of traffic due to accidents), medium-term changes (e.g., traffic rerouting during road construction) and long-term changes (e.g., after a shopping center has opened or closed). The predictive-channel-reservation (PCR) approach (Chiu and Bassiouni 2000) uses real-time position measurements to predict the future path of an MS. The utilization of real-time measurements introduces considerations not addressed in history-based schemes (e.g., the monitoring of changes in motion).

3.5.1.2 Channel reuse and interference adaptation

Based on the C/I constraint, we can view the cell served by a BS as the ensemble of points where the downlink and uplink C/I constraints are satisfied, i.e., the area where a good connection can be established. Traditionally, the interference factor has been modeled by a parameter called the *co-channel reuse distance*, which represents the minimum distance between two BSs sharing the use of a common channel. Fixed systems that do not employ adaptive designs are usually designed with safe worst-case-scenario margins of capacity and reuse distance, aimed at avoiding severe performance degradation (Chiu and Bassiouni 2000).

3.5.1.3 Predictive channel reservation

In the PCR scheme, channel allocation decisions are based on prediction (extrapolation) of the motion of MSs. Each MS periodically measures its position and orientation. The latter can be easily obtained from the vector of two consecutive position measurements. This information may be sent to the BS in uplink or may be readily available if the positioning is done by the BS itself. The BS uses the position/orientation information to predict the projected future path of the MS (Chiu and Bassiouni 2000). When a more advanced positioning technology (e.g., one that provides accurate estimation of velocity and acceleration) can be used, more complex extrapolation methods can then be used, similar to the methods developed for predicting the movement of simulated vehicles in networked virtual-reality training exercises (Bassiouni et al. 1997). Based on the projected path, the next cell that the mobile is moving towards is determined. When the MS is within a certain distance from that cell, the current BS sends a reservation request to the new BS in order to preallocate a channel and to accommodate the expected handoff event. A reservation may be deemed invalid (a false reservation) at a later time. In this case, a cancellation of the reservation must be sent to deallocate the reserved channel (Chiu and Bassiouni 2000).

3.5.2 Power control

The SINR is a limiting factor because of its direct relation to the variance of the TOA estimate. In practice, the SINR is increased using two techniques

(Prasad et al. 2000): (1) a technique based on the power-up phase in the uplink local-positioning (LP) scenario; and (2) another based on an idle slot in the downlink LP scenario. However, these techniques have the disadvantage of reducing the network capacity. To avoid this phenomenon, estimation methods resistant to the near–far problem could theoretically be applied directly to the received data to avoid the near–far problem. Unfortunately, in practice those methods are very expensive to implement (Muhammad et al. 2002).

The near–far problem happens when two transmitters, say A and B, one close to the receiver (e.g., A) and the other far away (e.g., B), transmit at the same time. As the useful transmitted signal of A is seen as noise by B, and vice versa, one can see that the SNR for B is much lower than that for A, assuming that the transmit powers of the two transmitters are comparable. This makes it more difficult, if not impossible, for transmitter B to be understood by the common receiver.

3.6 Cooperative Communications

In cooperative wireless communications, we are concerned with a wireless network where the users may increase their effective quality of service via cooperation. In a cooperative communication system, each user is assumed to transmit data and also to act as a cooperative agent for the other users (Figure 3.5).

Cooperation leads to trade-offs in terms of code rates and transmit power. In the case of the latter, on the one hand more power is needed because each user, when cooperating with other users, is transmitting for all of them. On the other hand, the baseline transmit power for all users will be reduced because of the beneficial effect of diversity. Out of this trade-off, one hopes to achieve a net reduction of the transmit power, assuming that everything else in the system is kept constant. A similar trade-off can be observed for the rate of the system. In a cooperative communication, each user transmits both their own bits and some information for a cooperative partner; this might cause a loss in the system rate. However, the spectral efficiency of each user improves because of the beneficial effect of cooperation diversity; as a consequence, the channel code rates can be increased. Again, a trade-off is observed. The premise of cooperation is that certain allocation strategies for the power and bandwidth of mobiles lead to significant gains in the system performance (Nosratinia et al. 2004).

3.6.1 RSS-based cooperative positioning

RSS data for both the BS–MS and the MS–MS links is available for cooperative positioning. Suppose MS_1 is the target MS, $\{X^{(i)} = [x^{(i)}, y^{(i)}]^{\mathrm{T}}, i = 1, \ldots, n_{\mathrm{ms}}\}$ are the coordinates of the MSs and $\{X^{[j]} = [x^{[j]}, y^{[j]}]^{\mathrm{T}}, j = 1, \ldots, n_{\mathrm{bs}}\}$ are the coordinates of the BSs. Similarly to the case of noncooperative positioning described in Section 3.2.5, Soork et al. (2008) have proposed a least-squares objective function

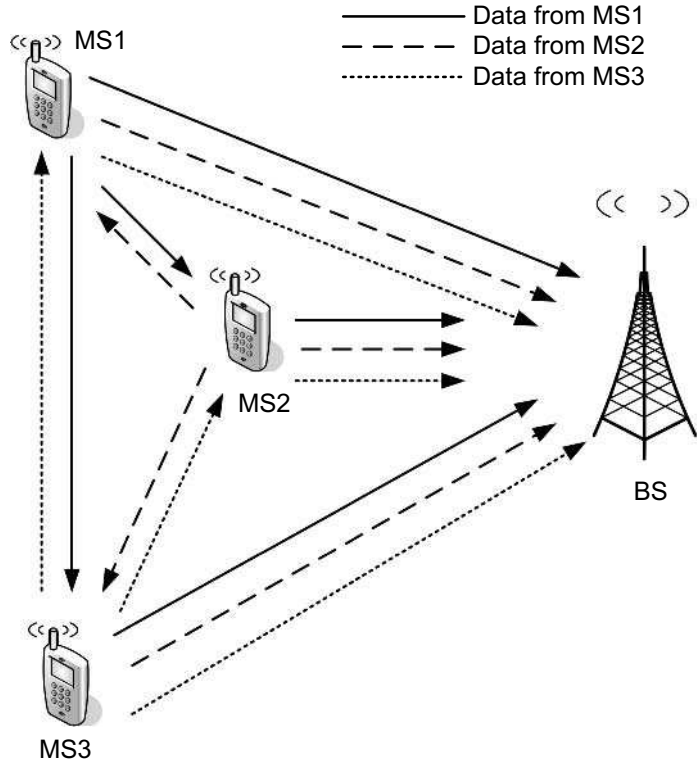


Figure 3.5 An example of a cooperative communication scenario.

as follows:

$$\mathcal{J}(X) = \sum_{i=1}^{n_{\text{ms}}} \sum_{j=1}^{n_{\text{bs}}} [p^{(i)[j]} - \alpha^{(i)[j]} + 10\beta \log(d^{(i)[j]})]^2 + \sum_{i=1}^{n_{\text{ms}}} \sum_{m=i+1}^{n_{\text{ms}}} [p^{(i)(m)} - \alpha^{(i)(m)} + 10\beta \log(d^{(i)(m)})]^2, \tag{3.24}$$

where $p^{(i)[j]}$, $\alpha^{(i)[j]}$ and $d^{(i)[j]}$ are, respectively, the received power, the α term and the distance for the link MS_i–BS_j. Similarly, $p^{(i)(m)}$, $\alpha^{(i)(m)}$ and $d^{(i)(m)}$ are, respectively, the received power, the α term and the distance for the cooperative link MS_i–MS_m.

The optimum point of $\mathcal{J}(X)$ in Equation (3.24), which is again obtained numerically, gives a position estimate for all MSs. As before, for further information about least-squares methods and the associated formalism, see Section 5.2.

3.7 Cognitive Radio and Mobile Positioning

Conventional static spectrum allocation results in low spectrum efficiency when we consider the continuously increasing demand for bandwidth. This has recently encouraged a lot of research in the field of cognitive radio (CR), which is seen as a promising technology for maximizing spectrum utilization. Fundamentally, CR tries to utilize radio resources by respecting the following rules: (1) sensing the surrounding spectrum environment, and (2) making intelligent decisions. Mitola and Maguire (1999) introduced the concept of the cognitive cycle, which consists of *radio scene sensing and analysis*, including *channel state estimation and prediction*, and *action* to transmit a signal. All actions in CR are based on the result of radio scene sensing and analysis, which analyzes RF stimuli from the radio environment, and then identifies the white spaces (available unused portions of the spectrum) and the position and transmission parameters of the various systems in use.

In the CR framework, there are two types of users: primary users, who have priority to access the spectrum, and secondary users, who have no priority or a lower priority than primary users. Hence, one of the challenging issues in CR is how to identify and analyze a primary user and the other systems that coexist with a CR system (Kim et al. 2008).

Kim et al. (2008) proposed a scheme to be used by the secondary users to estimate the unknown primary users' positions and transmission power, under the following assumptions (Figure 3.6). (1) The secondary users' positions are known. (2) The secondary users measure RSS values from primary users. (3) At least four secondary users receive a signal from each primary user (this assumption will be justified later, when we develop the equations of the model). (4) The shadowing effect is independent for each secondary user.

Let the real position of the primary user be $X = [x, y]^{\mathrm{T}}$ and the coordinates of the ith secondary user be $X_i = [x_i, y_i]^{\mathrm{T}}$, $i = 1, \ldots, n$, where n is the total number of secondary users receiving the primary user's signals. Let d_i denote the distance between the primary user and the ith secondary user:

$$d_i = \sqrt{(X - X_i)^{\mathrm{T}}(X - X_i)}, \quad i = 1, \ldots, n. \tag{3.25}$$

The ideal RSS or received power at the ith secondary user, denoted by p_i in watts, is expressed as

$$p_i = RSS^{(i)} = \phi_i \frac{p_i^{\mathrm{tx}}}{d_i^{\beta}}, \quad i = 1, \ldots, n, \tag{3.26}$$

where p_i^{tx} is the transmission power at the transmitter, ϕ_i represents all other factors that affect the received signal power, including antenna gain and antenna height, and β is the path loss exponent. We take the shadowing effect into account by using a log-normal path loss model:

$$p_i = RSS^{(i)} = \phi_i \frac{p_i^{\mathrm{tx}}}{d_i^{\beta}\eta_i^{\mathrm{shad}}}, \quad i = 1, \ldots, n, \tag{3.27}$$

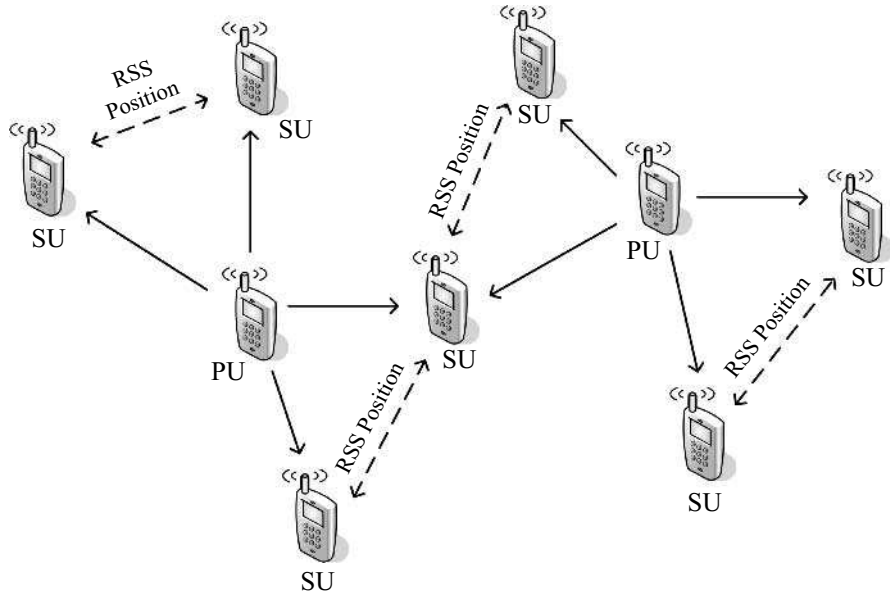


Figure 3.6 Cognitive radio and mobile positioning. "PU" stands for "primary user", and "SU" for "secondary user".

where $\eta_i^{\text{shad}} = 10^{0.1\eta_i}$ is a log-normal random variable and η_i is a Gaussian random variable with zero mean and variance σ^2 (Kim et al. 2008).

The RSS values at each user have severe disturbances caused by the shadowing effect; so, after measuring the raw RSS values, we have to refine the sample data by using the sample mean:

$$\overline{p_i}[\text{dBm}] = \frac{\sum_{k=1}^{m} RSS_k^{(i)}[\text{dBm}]}{m}, \tag{3.28}$$

where $RSS_k^{(i)}$ is the kth sample RSS value at the ith secondary user in dBm, and m is a predefined number of samples. Then, we can use a measurement model such as

$$r_i = \frac{\phi_i}{\overline{p_i}}, \tag{3.29}$$

where $\overline{p_i}$ is the sample mean of the RSS values at the ith node and ϕ_i is the propagation factor in Equations (3.26) and (3.27). The measurement model is based on the sample mean of the RSS values at each node. The above preprocesses reduce the shadowing disturbances in the RSS values in Equation (3.27).

When the number of samples for the measurement is m, the relationship between the measurement and the propagation model of Equation (3.27) is modeled as

$$r_i = \frac{\phi_i}{\overline{p_i}} \approx \frac{\phi_i}{RSS^{(i)}} = \frac{d_i^{\beta}}{p_i^{\text{tx}}}\eta_i^{\text{shad}}, \quad i = 1, \ldots, n, \tag{3.30}$$

where η_i^{shad} has the same meaning as in Equation (3.27) (Kim et al. 2008).

If there is no disturbance in the RSS measurement (no shadowing effect), Equation (3.27) can be expressed as

$$\frac{\phi_i}{RSS^{(i)}} = \frac{d_i^{\beta}}{p_i^{\text{tx}}} = \frac{[(x - x_i)^2 + (y - y_i)^2]^{\beta/2}}{p_i^{\text{tx}}}, \quad i = 1, \ldots, n. \tag{3.31}$$

By raising both sides of Equation (3.31) to the power of $2/\beta$, after some manipulations, we can obtain

$$x_i^2 + y_i^2 = 2xx_i + 2yy_i + \rho\left(\frac{\phi_i}{RSS^{(i)}}\right)^{2/\beta} - r, \quad i = 1, \ldots, n, \tag{3.32}$$

where $r = x^2 + y^2$ and $\rho = (p_i^{\text{tx}})^{2/\beta}$. Equation (3.32) holds for each secondary user that is receiving the primary user's signal; we can then express Equation (3.32) in matrix form,

$$H\mathsf{X} = B, \tag{3.33}$$

where

$$H = \begin{bmatrix} 2x_1 & 2y_1 & \left(\frac{\phi_1}{RSS^{(1)}}\right)^{2/\beta} & -1 \\ 2x_2 & 2y_2 & \left(\frac{\phi_2}{RSS^{(2)}}\right)^{2/\beta} & -1 \\ 2x_3 & 2y_3 & \left(\frac{\phi_3}{RSS^{(3)}}\right)^{2/\beta} & -1 \\ 2x_4 & 2y_4 & \left(\frac{\phi_4}{RSS^{(4)}}\right)^{2/\beta} & -1 \end{bmatrix}, \tag{3.34}$$

$$\mathsf{X} = \begin{bmatrix} X \\ \rho \\ r \end{bmatrix} = \begin{bmatrix} x \\ y \\ \rho \\ r \end{bmatrix} \tag{3.35}$$

and

$$B = \begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ x_3^2 + y_3^2 \\ x_4^2 + y_4^2 \end{bmatrix}. \tag{3.36}$$

As Equation (3.35) means that four parameters need to be estimated, we can justify the earlier assumption that at least four secondary users receive the signal from each primary user. The above estimation problem can be solved by using the LS method, with the aim of minimizing the disturbance caused by shadowing to the

measurements. Let $\hat{\mathbf{X}}$ be the estimated position; then the solution can be computed as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}}\{(H\mathbf{X} - B)^{\mathrm{T}}(H\mathbf{X} - B)\} \tag{3.37}$$

$$= (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}B. \tag{3.38}$$

In Kim et al. (2008), a constrained, weighted optimization method to solve Equation (3.33) taking into account disturbances in measurements was also proposed.

3.8 Conclusions

In this chapter, the basics of wireless communications and how those are related to mobile positioning have been discussed. The radio propagation scenario has been introduced and the various physical phenomena that underlie wireless communications have been classified. We have presented the various multiple-antenna techniques and some considerations of the possibilities offered by such techniques in terms of mobile positioning. Modulation and multiple-access techniques considered relevant for future wireless positioning systems have been reviewed. Some RRM techniques have been presented, and we have outlined how localization can benefit from them, and vice versa. Finally, two of the emerging areas in wireless communications, i.e., cooperative communications and cognitive radio technology, have been discussed, and their relation to mobile positioning has been highlighted.

4

Fundamentals of Positioning

4.1 Introduction

Whereas previous chapters have presented position-related aspects of wireless services and wireless communications, the current chapter will focus on positioning methods. Positioning in wireless networks is possible by exploiting the characteristics of the propagation signals. Certain characteristics of the signal must be measured, such as the received power, in order to permit the estimation of position. The strict requirement is that the measured information has to have a physical relation to the position of the device; otherwise, the information is irrelevant for positioning purposes. The relation between the measurements and the position is always corrupted by a noise component, adding unavoidable complexity to a positioning system. For this reason, it is necessary to have mathematical mechanisms that are able to manipulate the measurements in order to extract position coordinates. The more advanced methods include a statistical treatment of the errors in the measurements.

This chapter will start by describing the types of measurements that are commonly possible in typical wireless communication systems, and then it will introduce some techniques for estimating position based on those wireless measurements. The chapter continues by describing some of the errors that exist in the measurements and presenting some metrics of accuracy commonly used within the context of positioning.

4.2 Classification of Positioning Infrastructures

Positioning systems can be classified according to several different key characteristics. The typical classification concerns the topology of the system, in particular the methodology used to obtain measurements and subsequently process them in order to determine location information. Some alternative classifications are, for instance, related to the properties of the communication link or to the *type of*

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

integration into the communication technology. In the classification based on the communication link, systems are grouped according to their similarities with respect to the propagation conditions. The classification according to the type of integration groups systems depending on the approach used to integrate positioning systems and solutions into the communication technology.

4.2.1 Positioning-system topology

As defined by Drane et al. (1998) and Vossiek et al. (2003), the classification of systems with respect to system topology groups systems according to where measurements are made and where those measurements are processed. Figure 4.1 shows this classification. In general, the entity responsible for the measurements and the entity responsible for its subsequent processing are not necessarily the same. When the measurements are made and processed in the MS, the system is classified as a self-positioning or mobile-based positioning system (Figure 4.1(a)). In order to allow the MS to self-position itself, it is necessary to inform the MS about the location of the BS. This location can either be widely known in the network or be sent by each BS during the MS–BS communication. The major advantage of this solution is that it is scalable and simple, but on the other hand it does not allow tracking of the MS from the point of view of the network. When measurements are both obtained and subsequently processed on the network side, the solution is referred to as remote positioning or network-based positioning (Figure 4.1(b)). The system works in such a way that signals emitted by the MS or reflected from it are received by the network BSs. In this case the position solution suffers from scalability problems, because the network may be overloaded with signaling and processing when the number of MSs increase. When measurements are obtained in one entity but processed in another, the systems are said to be "indirect", either an indirect self-positioning system (Figure 4.1(c)) if the network measures the signals and the MS processes them, or an indirect remote positioning system (Figure 4.1(d)) in the other case. These solutions are also referred to as mobile-assisted positioning. Owing to the differentiation between the entities that measure and the entities that process information, it is necessary to establish a temporary link between the MS and the BS so that information can be exchanged.

Table 4.1 shows a summarized description of this classification. Despite the class of the system, position information may be needed in any element of the network, meaning that it may be necessary to establish additional links after the estimation of position so that the information is exchanged.

4.2.2 Physical coverage range

Another type of classification concerns the physical coverage range of the communication technology. This classification groups systems according to the properties of the propagation conditions of the signals. Thus, the classification is generally into satellite, cellular and short-range positioning. Table 4.2 shows

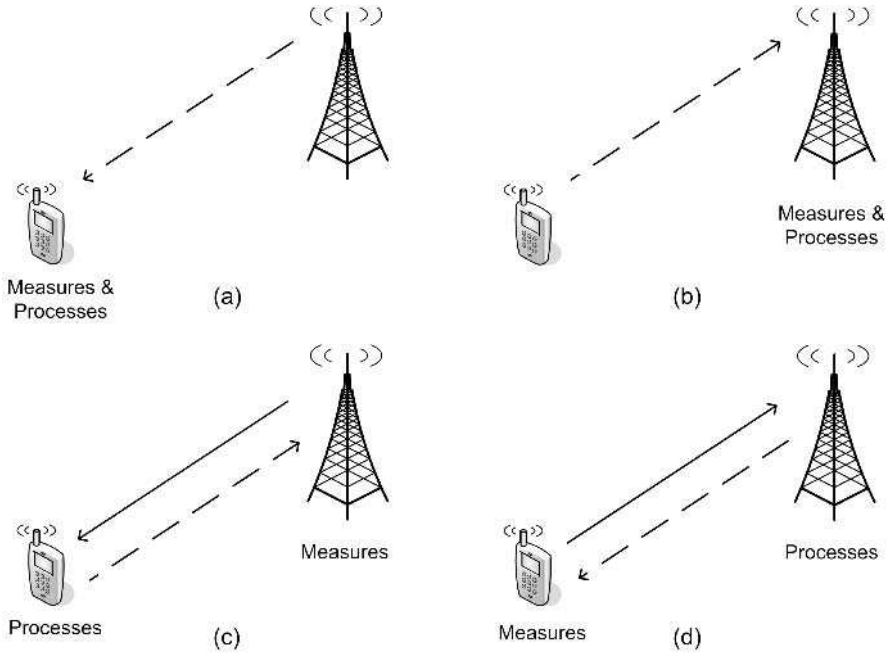


Figure 4.1 Classification according to system topology: (a) self-positioning; (b) remote positioning; (c) indirect self-positioning; (d) indirect remote positioning. Dashed lines represent a communication signal used for measuring the wireless channel, and solid lines represent an actual transfer of measured data.

Table 4.1 Classification of positioning solutions based on system topology

Concept	**Description**
Self-positioning	Measurement and processing done by the MS
Remote positioning	Measurements and processing done by the network
Indirect self-positioning	Measurements done by the network and processing done by the MS
Indirect remote positioning	Measurements done by the MS and processing done by the network

an overview of each solution. Satellite solutions are based on space–earth communication. For this reason, problems such as propagation in the ionosphere are to be considered. Regarding cellular solutions, the most important problems concern NLOS, scattering, shadowing and multipath effects. For short-range positioning, the propagation statistics are considerably dependent on the scenario.

Table 4.2 Classification of positioning solutions based on physical coverage range

Concept	Description
Satellite	Positioning solution requires satellite communication
Cellular	Cellular networks are used for obtaining position information
Short-range	Positioning in WLAN/WPAN networks

Satellite positioning systems, such as the GPS, are aimed at providing positioning information on a global scale. The satellites, orbiting around the Earth, have a communication range *of virtually everywhere*. Of course, it is necessary to take into account the limitations of such systems, such as their inability to position users when they are indoors. Cellular positioning concerns methods for positioning MSs within the range of technologies such as GSM and UMTS (see Chapter 8). For these systems, the communication is limited by the range of the cells. The propagation conditions are commonly urban scenarios. The short-range positioning solutions run over WLAN/WPAN technologies such as IEEE 802.11 or Bluetooth. For short-range communications, the typical scenario is indoors.

In this classification, the RFID technology can be considered as an extra case, since propagation conditions are rarely treated in the context of positioning information (Section 8.4.2.1).

4.2.3 Integration of positioning solutions

As suggested by Figueiras (2008), positioning solutions can also be classified according to the way that the positioning system is integrated into the communication technology. Systems can be classified, according to Table 4.3, into integrated, opportunistic and hybrid systems. The integrated solutions relate to stand alone positioning systems that are adapted to interact with communication solutions. The current integration of GPS receivers into smart phones is a typical case of this class of solutions. The second group, the opportunistic solutions, are generally implemented in software. These solutions exploit the properties of the communication technology in order to estimate the user's position. In communication networks making use of power control techniques, these solutions can be easily applied, because the technology itself has already implemented a mechanism for measuring signal strength. The price to pay is that the opportunistic solutions are commonly less accurate than integrated solutions. Finally, the hybrid solutions make use of both integrated and opportunistic solutions in the same system. These hybrid solutions are appropriate for combining results or for adapting the positioning scheme according to the most accurate solution given the conditions.

Table 4.3 Classification of positioning solutions based on the type of integration

Concept	Description
Integrated	The positioning system is integrated with the wireless communication technology
Opportunistic	The available communication links are used opportunistically for estimating location information
Hybrid	Both of the above types are used in a cooperative fashion or alternately, depending on the scenario/conditions

4.3 Types of Measurements and Methods for their Estimation

Despite the large number of different positioning systems and solutions for wireless networks, the types of measurements used to determine user positions are rather few. The measurements can be base station-dependent identity labels and/or link-dependent properties such as RSS. Whatever their type, these measurements are the output of the measurement layer defined by the Location Stack in Section 2.2.

4.3.1 Cell ID

Cell identification, also called cell ID (or CID), is a technique for localization which estimates the location of the MSs as the position of the fixed reference points (FRPs) which the MSs are connected to. In practice, "cell ID" means an actual measurement obtained in the positioning system. This cell ID is then indexed in a database in order to determine the position and the accuracy range of the FRP. The accuracy of this technique is therefore limited by the physical communication range of the FRPs. While in short-range networks the physical range is some tens of meters, in cellular networks the range can be some few thousands of meters. Thus, if no additional information is available, the accuracy of the system is given by the communication range of the wireless technology. Also, in order to perform unambiguous positioning, the FRPs need to be uniquely identified in the whole network. For instance, in a WLAN or Bluetooth network, the MAC address is a commonly used identifier. In cellular networks, FRPs can be identified by their mobile country code (MCC), mobile network code (MNC), local area code (LAC) and cell ID. In theory, for WLAN networks, the MAC address is virtually unique (IEEE 2002), but in practice the address can be programmed and changed by software (kingpin 1998). Regarding the identifier in cellular networks, it is common that not all of the required fields are available. This can introduce some indecision into the identification of the FRP. Additionally, it is common that FRPs in cellular systems have several (commonly three) sections, which increases the complexity of the system.

Cell ID is often used in combination with other types of information, such as range and/or angle measurements, coverage map information or cartographic mapping (Gustafsson and Gunnarsson 2005b). One of the major problems with such information is that the positions of the FRPs are typically treated as confidential by the network operators. This implies that such kinds of solutions must often be used in an opportunistic fashion, where the localization service is provided by third-party entities.

4.3.2 Signal strength

The RSS is a measure of the magnitude of the power of the signal that the receiver sees at its terminals. Owing to the destructive propagation effects presented in Section 3.2, this measure is typically subject to a noise component with variations of tens of dB for the typical scenarios of interest, i.e., indoor short-range localization and outdoor long-range localization.

The RSS can be modeled by a constant component and a variable component. Generally, the constant component is modeled by a path loss propagation model and the variable component is modeled by a range of complex propagation effects (see Section 3.2). These propagation effects can be, for instance, signal attenuation, shadowing, multipath effects, scattering and diffraction. For an overview of these propagation effects see Section 3.2, and for a deeper insight see, for example, Parsons (2000) and Goldsmith (2005).

To model the path loss effect, the following equation is considered:

$$p = \alpha - 10\beta \log(d), \tag{4.1}$$

where p is the received signal strength; α is a parameter dependent on the transmission power, additional losses and other system-component-dependent constants; β is the path loss exponent; and d is the physical length of the communication link, i.e., the distance between transmitter and receiver.

Regarding the variable component, it is complex to obtain a model that takes into account the stochastic behavior of the propagation effects involved. The reason is that these effects strongly depend on the scenario. Thus, the most generic approach is to consider all these effects as a single additive Gaussian-distributed noise component with a standard deviation of σ_{shad}:

$$p = \alpha - 10\beta \log(d) + \eta_{\text{shad}}, \quad \eta_{\text{shad}} \sim \text{Norm}(0, \sigma_{\text{shad}}). \tag{4.2}$$

Although this noise component is generally considered as Gaussian throughout the present section, more detailed models are considered in Section 4.5.1.

4.3.3 Time of arrival

The TOA is the timestamp that a receiver sees on its internal clock when a signal is received at its terminals. With this type of measurement, one of the factors that

most influences the error that corrupts TOA observations is the granularity of the internal clock. Given that the clock increments only in the instants when a clock tick occurs, any signal trigger between two consecutive clock ticks will result in the same timestamp reading. Thus, for every instant within this period, the estimated link length will be the same. In terms of distance, the maximum error due to clock granularity can be easily calculated as the quotient of the speed of light c and the clock frequency f:

$$\varepsilon_{\text{clock}} = \frac{c}{f}. \tag{4.3}$$

When this error has the same magnitude as the typical length of the communication channel, it becomes unfeasible to estimate the link length based on TOA measurements. Although the TOA is strongly dependent on the technology, it is common to consider it, in general, as modeled by the sum of a constant component and a variable noise component.

The constant component t of the TOA model is equivalent to the sum of the transmission time t_{tx} (the timestamp in the transmitter clock when the signal is transmitted) and the propagation delay (the time that the signal needs to travel from the transmitter to the receiver):

$$t = t_{\text{tx}} + \frac{d}{c} + \varepsilon_{\text{sync}}. \tag{4.4}$$

In Equation (4.4), c is the propagation velocity in the medium, commonly the speed of light, and d the physical length of the communication link. The component $\varepsilon_{\text{sync}}$ is a deterministic quantity that compensates any mis-synchronization between the transmitter and the receiver. Note that Equation (4.4) requires the transmitter and the receiver to be either clock synchronized (where $\varepsilon_{\text{sync}} = 0$) or nonsynchronized but with a known difference (where $\varepsilon_{\text{sync}} \neq 0$), otherwise the transmitted timestamp would not have the same reference as the received timestamp. A major complexity related to TOA is knowing $\varepsilon_{\text{sync}}$, which often is not possible to obtain. This can represent a strong disadvantage of TOA observations because clock-synchronizing a whole network is commonly complex to implement and expensive to maintain.

Regarding the variable component, it is possible to divide the problem into clock granularity and propagation effects. While the clock granularity introduces a uniform error between 0 and the value given by Equation (4.3), the propagation effects require a complex treatment, owing to the unpredictability of the impact that such effects have on communications. In other words, the propagation effects depend strongly on the scenario considered. For this reason, the error due to these effects can be generalized and simplified into a Gaussian distribution:

$$t = t_{\text{tx}} + \frac{d}{c} + \varepsilon_{\text{sync}} + \eta_{\text{clock}} + \eta_{\text{prop}}, \quad \begin{cases} \eta_{\text{clock}} \sim \text{Unif}(0,\, c/f), \\ \eta_{\text{prop}} \sim \text{Norm}(0,\, \sigma_{\text{prop}}), \end{cases} \tag{4.5}$$

where σ_{prop} represents the standard deviation of the errors introduced by the propagation effects. The value of σ_{prop} is commonly taken from the literature or

obtained by a precalibration phase. Note that in Equation (4.5), to assume that $\varepsilon_{\text{sync}}$ is known is mathematically equivalent to assume that the clocks are synchronized between transmitter and receiver.

4.3.4 Time difference of arrival

The TDOA is the difference between two TOA measurements obtained from two equivalent signals emitted at exactly the same time. This technique was proposed as a solution that allows one to drop the requirement that TOA measurements impose clock synchronization between transmitter and receiver. Since the TDOA is the difference between two equivalent signals sent at the same time, the dependency shown in Equation (4.4) on the time of transmission of the signal is lost. Thus, assuming that at time t_{tx} two equivalent signals are sent from BS_1 and BS_2, and given that the MS is d_1 meters from BS_1 and d_2 meters from BS_2, the TDOA t seen at the MS is given by:

$$\mathsf{t} = t_{\text{tx}} + \frac{d_1}{c} + \varepsilon_{\text{sync}} - \left(t_{\text{tx}} + \frac{d_2}{c} + \varepsilon_{\text{sync}} \right) = \frac{d_1 - d_2}{c}. \tag{4.6}$$

This technique assumes two possible modes, uplink and downlink. In the uplink mode, the MS produces a signal which is received at two different BSs. This mode may have scalability problems, since all the operations are performed in the network. In contrast, in the downlink mode, two BSs transmit, at the same time, two equivalent signals, which are received by the MS. This mode is more scalable than the uplink mode since calculations are performed in the MS. Whatever the mode of operation is, the BSs need to be clock synchronized, though BSs and MSs do not need to be reciprocally synchronized.

Introducing a variable component into the TDOA is straightforward if we use Equation (4.5):

$$\mathsf{t} = \frac{d_1 - d_2}{c} + \eta_{\text{clock}} + \eta_{\text{prop}}, \quad \begin{cases} \eta_{\text{clock}} \sim \text{Tri}(-c/f,\, c/f), \\ \eta_{\text{prop}} \sim \text{Norm}(0,\, 2\sigma_{\text{prop}}). \end{cases} \tag{4.7}$$

In Equation (4.7), the subtraction of two uniform distributions results in a triangular distribution between $-c/f$ and c/f centered at 0. Similarly, the subtraction of two equal normal distributions results in a normal distribution with variance $2\sigma_{\text{prop}}$ and centered at 0. In Equation (4.7), it is assumed that the noise components are equal in the communication channels between the MS and BS_1 and between the MS and BS_2. The extension to unequal distributions is straightforward.

4.3.5 Angle of arrival

The AOA is a measure of the angle at which the signal arrives at the receiver. The technical procedure for obtaining AOA measurements requires directional antennas, and for this reason its implementation may represent additional cost and system

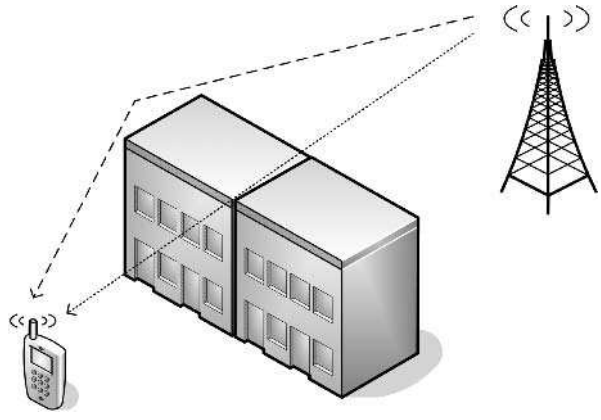


Figure 4.2 Effect of shadowing on AOA measurements. The dotted line represents the direct angle between transmitter and receiver and the dashed line represents the propagation path influenced by the shadowing effect.

complexity. In general, the AOA is related to an internal reference, which is commonly chosen based on the topology of the network. For instance, in multi-BS scenarios, the angles are commonly referenced to a virtual line that connects two BSs. A major problem with this type of measurement is the shadowing effect (see Section 3.2.2) present in the propagation channel. Since the shadowing effect biases the trajectory described by the main signal component, the signal received will arrive from an angle different from that described by the physical positions of the transmitter and the receiver. Figure 4.2 shows this problem. Owing to the complexity or even impossibility of detecting the influence of this effect, it is acceptable to use an estimate of the angle as a rough indication of position in terrestrial systems.

A simple, general model of the AOA is to assume an internal reference which all angles and coordinates are related to, i.e., all angles and the entire coordinate system are relative to that reference. Then, assuming that the receiver is placed at position (x, y) and the transmitter at position $(x_{\text{tx}}, y_{\text{tx}})$, the AOA θ can be modeled by:

$$\theta = \text{atan}\left(\frac{y_{\text{tx}} - y}{x_{\text{tx}} - x}\right). \tag{4.8}$$

Note that the model of Equation (4.8) does not give a unique solution for (x, y). For this reason, it is necessary to implement angulation techniques for estimating a single position of the MS; these are surveyed in Section 4.4.2.3.

Regarding the noise component of AOA measurements, Goldsmith (2005) defines the noise as Gaussian. Thus, by assuming that this component has a zero-mean Gaussian distribution with standard deviation η_{aoa}, the model of the AOA can be written as:

$$\theta = \text{atan}\left(\frac{y_{\text{tx}} - y}{x_{\text{tx}} - x}\right) + \eta_{\text{aoa}}, \quad \eta_{\text{aoa}} \sim \text{Norm}(0, \sigma_{\text{aoa}}). \tag{4.9}$$

4.3.6 Personal-information identification

Personal-information identification is based on another type of data, not directly related to wireless positioning, but is of wide use, especially in positioning for forensic investigation (Mohay et al. 2003). The type of data used can be data from airplane tickets, government census data, car identification information, credit card data, data from security cameras, computer-based information (such as wireless MAC addresses, login user names or Internet Protocol (IP) addresses), person-dependent characteristics (such as fingerprints, eye data or DNA data) or words of other people, among many other sources of data. This type of positioning typically relies on robust data-mining techniques capable of extracting meaningful relationships that can be used to estimate or predict movement trends and patterns both in the past and in the future. This topic will not be further discussed, since it is not the main focus of this book.

4.4 Positioning Techniques

Based on the various measurements obtained in a wireless communication, as reviewed in the previous section, the current section aims introducing the algorithms and methods used in order to estimate the position of an MS. The methods presented in this section are the basic methods used in positioning applications. More complex methods are introduced in later chapters.

4.4.1 Proximity sensing

Proximity-sensing techniques are techniques that do not use the range or angles of communication links. Instead, they identify the mobile user's position once the user is within the physical communication range of a device.

4.4.1.1 Physical contact

This is the simplest technique, which localizes the MS when it establishes physical contact with the BS. Here, the use of the terms "MS" and "BS" may be somewhat inaccurate, since this techniques relies basically on sensors of pressure, contact or distance. If the positions of the sensors are known, it is possible to localize the user once a contact is established.

4.4.1.2 Identity methods

Identity methods are methods that rely on the position of a reference point in order to estimate the position of the MS. This technique is common in RFID, but also widely used in all types of cellular system. Figure 4.3 shows a representation of this technique. Once the MS enters the range of a specific BS, the cell ID of the BS is read and is subsequently used to index a database that returns the position of the BS and possibly its communication range. The position of the BS is then assumed to be the

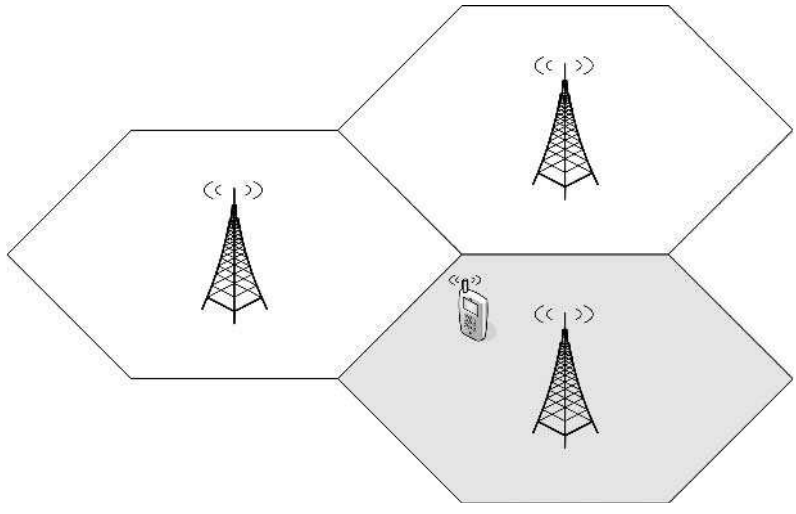


Figure 4.3 Cell ID technique.

position of the MS given that no other information is available. The communication range of the BS is assumed to be the best estimator for the accuracy of the position estimator. Section 8.3.2.1 introduces some more advanced techniques.

Identity methods are the simplest positioning techniques, but, at the same time, they are the least accurate methods for wireless positioning. A typical use of this technique is to support other positioning methods in hybrid approaches or establish geometrical constraints on the estimators.

A simple algorithm capable of estimating the position of MSs based on the locations of all the BSs that each MS can communicate with is the centroid method (Cheng et al. 2005). This algorithm calculates an arithmetic mean of the coordinates of all the BSs detected by the MS. Thus, assuming that the MS detects n BSs located at positions $\{X_i = [x_i, y_i]^{\mathrm{T}}, i = 1, \ldots, n\}$, the estimated position $X = [x, y]^{\mathrm{T}}$ of the MS is:

$$X = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{4.10}$$

As an alternative to the case when more information is available, the centroid algorithm can be weighted either by a measurement of range or by the actual communication range of each BS:

$$X = \frac{\sum_{i=1}^{n} (d_i^{\max})^{-1} X_i}{\sum_{i=1}^{n} (d_i^{\max})^{-1}}. \tag{4.11}$$

4.4.1.3 Macropositioning

Macropositioning concerns data-mining techniques commonly used together with personal-information identification. These methods are not directly related to

wireless positioning, but represent an important group of methods in the wide field of positioning.

4.4.2 Triangulation

In contrast to the identification methods mentioned above in Section 4.4.1.2, triangulation methods use range or angle measurements in order to estimate the position of the MS. These methods use trigonometry in order to combine data from several sources. Generally, these type of methods are supported by identity methods, since range and angle measurements require the position of the BSs to be known.

4.4.2.1 Lateration

Lateration is a technique that performs localization by using the distances between BSs and MSs. The calculation of the predicted location can be seen as the problem of determining the third vertex of a triangle given the lengths of the three edges and the location of two vertices. As Figure 4.4 shows, the two known vertices of the triangle are given by the positions of two BSs with known locations, one of the edges is given by the distance between the two BSs and the remaining two edges are given by the estimated length of the MS–BS communication channel. The estimated length can be calculated based on range observations such as RSS (see Section 4.3.2) or TOA (see Section 4.3.3) measurements. Although two BSs define the concept of triangulation, there is still a "flipping" problem around the axis connecting the two BSs. For this reason, with only two BSs, the predicted location may result in two points for a two-dimensional (2D) space. Thus, the minimum number of BSs necessary is three (placed at noncollinear points), where the third BS helps in determining which of the two positions is more probable. A similar idea can be used to argue that four noncoplanar points are the minimum number of BSs for three-dimensional (3D) localization.

As Figure 4.4 shows, the triangulation method requires in a 2D scenario three BSs with known locations. However, owing to the noisy behavior of the measurements, it is not possible to determine a single estimate of the MS position using trigonometric manipulation. The problem is that the link length estimators are corrupted by noise (see Section 4.3) and consequently they do not intersect in a single location. Thus, in order to estimate the position of the MS, it is necessary to combine data from the three BSs. The straightforward solution, although not the most elegant, would be to use two BSs to determine the intersection points and then use the third BS to choose one of these two points. This approach has two problems: firstly, it does not use data from the third BS, except for using it as a decision criterion; and, secondly, it does not handle the case when the circles do not intersect. A more elegant solution can be obtained as follows. Let us assume that the distance between the MS, located at position $X = [x,\ y]^{\mathrm{T}}$, and BS_i, located at $\{X_i = [x_i,\ y_i]^{\mathrm{T}}\}$, is given by d_i, where:

$$d_i^2 = (x_i - x)^2 + (y_i - y)^2. \tag{4.12}$$

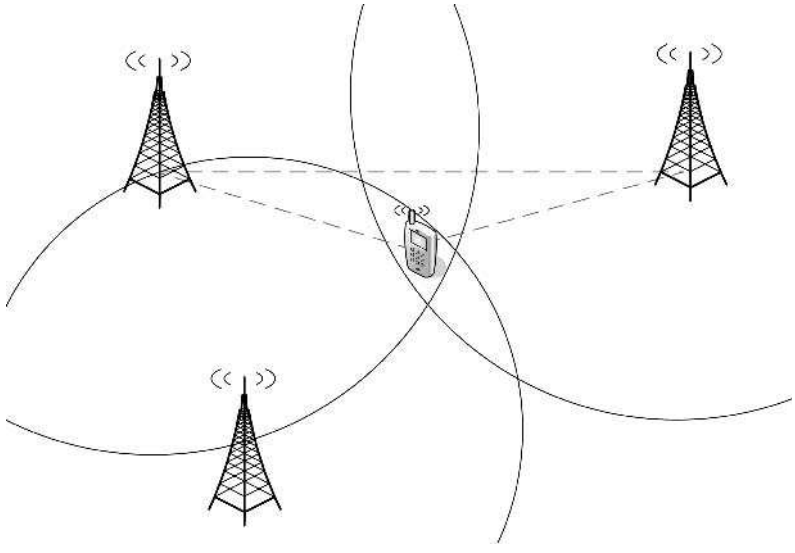


Figure 4.4 Triangulation technique.

Then, by subtracting Equation (4.12) for BS_i from the equivalent expression for BS_1, it is possible to obtain:

$$d_i^2 - d_1^2 = x_i^2 + y_i^2 - x_1^2 - y_1^2 - 2x(x_i - x_1) - 2y(y_i - y_1). \tag{4.13}$$

We can write this in matrix notation as:

$$HX = B, \tag{4.14}$$

where

$$H = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_n - x_1 & y_n - y_1 \end{bmatrix}, \tag{4.15}$$

and

$$B = \frac{1}{2} \begin{bmatrix} (d_1^2 - d_2^2) + (x_2^2 + y_2^2) - (x_1^2 + y_1^2) \\ (d_1^2 - d_3^2) + (x_3^2 + y_3^2) - (x_1^2 + y_1^2) \\ \vdots \\ (d_1^2 - d_n^2) + (x_n^2 + y_n^2) - (x_1^2 + y_1^2) \end{bmatrix}. \tag{4.16}$$

According to the least-squares algorithm formalized in Section 5.2, the solution for Equation (4.14) is given by:

$$X = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} B. \tag{4.17}$$

For more advanced algorithms, see Chapter 5 or Kailath et al. (2000) and Sayed and Yousef (2003).

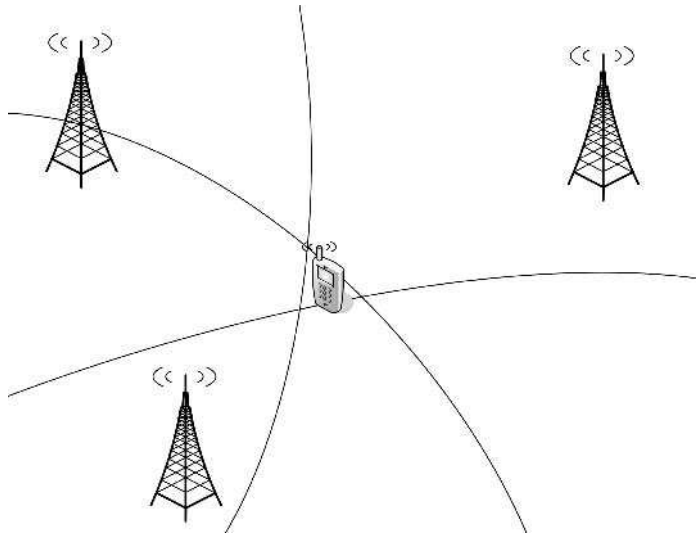


Figure 4.5 Hyperbolic localization technique.

4.4.2.2 Hyperbolic localization

Hyperbolic localization is an alternative to the triangulation technique when link lengths are estimated using TDOA measurements. Since these types of measurements do not depend on the link length but on the difference between two measures of link lengths, the TDOA measurements describe a hyperbola as the possible location of the user position (see Figure 4.5). For a 2D space, it is possible to estimate the (x, y) position of the MS from a system of equations with three different TDOA readings. For the same reason for which we need a third BS in order to obtain a unique solution using triangulation, hyperbolic localization requires a fourth BS.

Similarly to the lateration technique, hyperbolic localization is subject to the problem of non-intersecting curves. The same approaches as those used to solve the problem in the case of triangulation can be used in this case. Thus, assuming that TDOA measurements t_i result in a difference of distances d_i between BS_i and BS_1[1], the problem can be solved as follows:

$$\mathsf{d}_i = d_i - d_1, \tag{4.18}$$

$$(\mathsf{d}_i + d_1)^2 = d_i^2, \tag{4.19}$$

$$\mathsf{d}_i^2 + 2\mathsf{d}_i d_1 = x_i^2 + y_i^2 - x_1^2 - y_1^2 - 2x(x_i - x_1) - 2y(y_i - y_1). \tag{4.20}$$

We can write this in matrix notation as:

$$H\mathsf{X} = B, \tag{4.21}$$

[1] Referenced to BS_1.

where

$$\mathsf{X} = [X^{\mathrm{T}}, d_1]^{\mathrm{T}}, \tag{4.22}$$

$$H = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 & \mathsf{d}_2 \\ x_3 - x_1 & y_3 - y_1 & \mathsf{d}_3 \\ \vdots & \vdots & \vdots \\ x_n - x_1 & y_n - y_1 & \mathsf{d}_n \end{bmatrix}, \tag{4.23}$$

$$B = \frac{1}{2} \begin{bmatrix} (x_2^2 + y_2^2) - (x_1^2 + y_1^2) - \mathsf{d}_2^2 \\ (x_3^2 + y_3^2) - (x_1^2 + y_1^2) - \mathsf{d}_3^2 \\ \vdots \\ (x_n^2 + y_n^2) - (x_1^2 + y_1^2) - \mathsf{d}_n^2 \end{bmatrix}. \tag{4.24}$$

The solution is then given by:

$$\mathsf{X} = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} B. \tag{4.25}$$

Note that in Equation (4.21), the distance between the MS and BS_1, d_1, is included in the vector X on the left because it is itself also unknown.

4.4.2.3 Angulation

Angulation is a technique that makes use of measurements of AOA. The algorithm relies in a 2D scenario on two BSs at known locations capable of measuring AOA (see Figure 4.6). In a 3D space, it is possible to use this technique if another measurement of azimuth is available. In this technique, a possible convention is to predefine the direction of one of the vectors that connects two BSs as the magnetic north. This vector is defined as the reference at 0°. Consequently, the magnetic south is defined as 180°. Each BS must be able to identify the angle that the transmitted signal originated from, referenced to the predefined magnetic north. Several methods can be used to measure this angle, such as the use of directive antennas (Vanderveen et al. 1997) or methods making use of the phase properties of the signal (Departments of Defense and of Transportation 2001). After measuring the angles and knowing the location of the BSs, it is possible to predict the location of the MS.

Due to the fact that angulation uses AOA, the problem of systematic errors introduced by shadowing is present here. Since this problem may be not noticeable by the BSs, the errors may be prohibitive for pinpointing a position accurately. However, for a rough indication of the area where the MS is placed, angulation can be a good candidate solution for localization.

Assuming that each BS i is capable of measuring the AOA θ_i of the signal transmitted from the MS, based on Equation (4.8), it is possible to obtain:

$$(x_i - x)\sin(\theta_i) = (y_i - y)\cos(\theta_i), \tag{4.26}$$

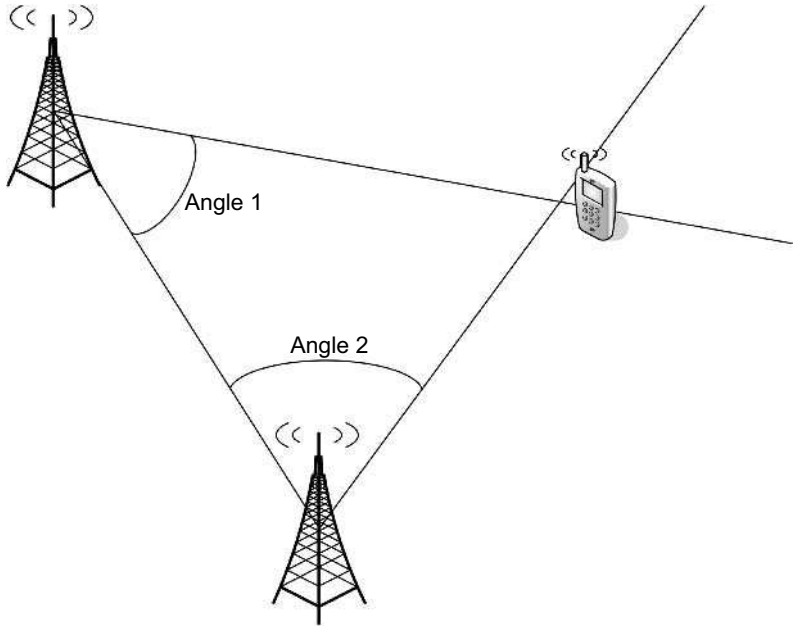


Figure 4.6 Angulation technique.

and in matrix form:

$$HX = B, \tag{4.27}$$

where

$$H = \begin{bmatrix} -\sin(\theta_1) & \cos(\theta_1) \\ -\sin(\theta_2) & \cos(\theta_2) \\ \vdots & \vdots \\ -\sin(\theta_n) & \cos(\theta_n) \end{bmatrix}, \tag{4.28}$$

and

$$B = \begin{bmatrix} y_1 \cos(\theta_1) - x_1 \sin(\theta_1) \\ y_2 \cos(\theta_2) - x_2 \sin(\theta_2) \\ \vdots \\ y_n \cos(\theta_n) - x_n \sin(\theta_n) \end{bmatrix}. \tag{4.29}$$

The solution is given by:

$$X = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} B. \tag{4.30}$$

4.4.3 Fingerprinting

This group of techniques performs localization of MSs based on information obtained from the surroundings of the MSs. For these techniques, it is important to have good knowledge about the surroundings in order to correlate the information obtained

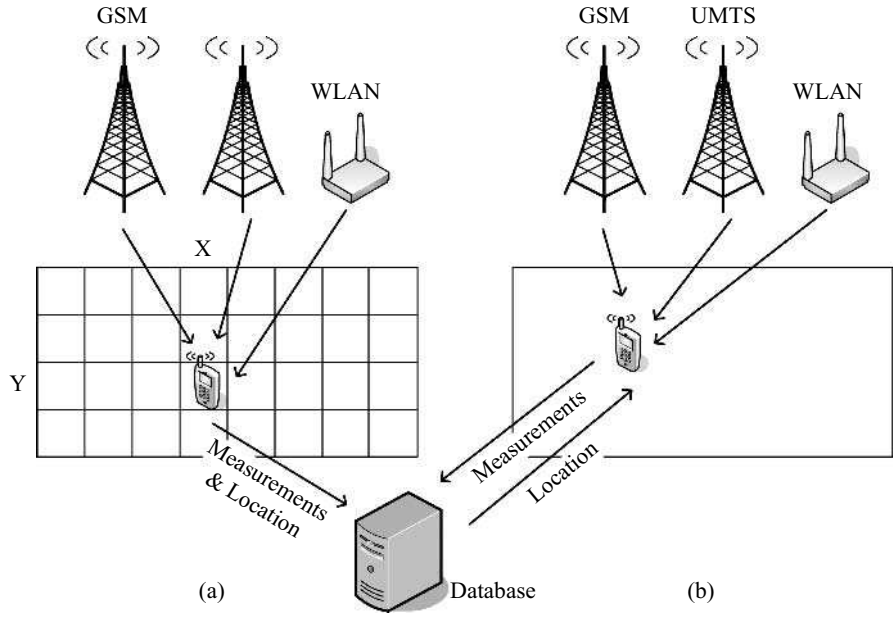


Figure 4.7 Database correlation technique: (a) calibration procedure; (b) localization procedure.

during the location process. This requires effort in the calibration of the system, which may not be desirable or not even implementable for some specific applications.

In this technique, the accuracy localizing the MS depends strongly on the information contained in the database. This database must be as complete as possible, with (if possible) as much variety of measurements as possible, such as received power, time of flight and angle of arrival. As we can see in Figure 4.7, it is also possible to consider measurements from several systems such as GSM, UMTS systems or any other technology enabled into the MS.

The greater the amount of information in the database, the more complex are the algorithms for localizing MSs and also the greater is the complexity of managing the database. Fingerprinting techniques are very dependent on the scenario. If something that may influence the propagation conditions changes, for example, a door opening in an indoor scenario, the database must be recalibrated all over again.

As we can see in Figure 4.7, during the calibration phase the measurements must be performed over the possible area with a certain step size. The locations and measurements are then passed into the database for further processing. During the localization phase, the MS must be able to measure some of the possible signals and relay them to a central server. This server, after correlating these measurements with the information included in the database, will then determine the predicted location. Section 5.5.1 presents an example of this technique.

4.4.3.1 Calibration phase for database creation

In general, fingerprinting techniques are of high accuracy; however, it is necessary to have a database with as much information as possible. Obtaining this information can be a major problem and virtually impossible for global positioning. In small areas, the creation of the database can be done by dividing the entire area into small clusters. Then, in each cluster, measurement samples are obtained and stored in the database. The smaller the clusters, the higher the accuracy is. This precalibration phase is commonly cumbersome and time-consuming, and becomes unfeasible for some applications such as positioning in large areas. In the extreme case of global positioning, it is not possible to cover the whole world. In this case, the approach of creating the database is different; the database is often composed of information which identifies the FRPs. In this case, the methods used are commonly not ambitious enough to aim at characterizing subcell-scale areas.

Besides the difficult phase of creating the database, there is a problem related to the maintenance of the database itself. If the propagation conditions change during the operations of the system, a recalibration phase needs to be run. This problem makes it considerably harder to use such fingerprinting positioning systems. In the particular case of global positioning, possible solutions to this problem of database maintenance may rely on large-scale collaborative schemes.

4.4.3.2 Image/video approaches

These techniques use video or image techniques in order to locate devices. In order to localize an MS, the surroundings of it are filmed or photographed, and analyzed. Based on a priori knowledge of the environment or robust artificial intelligence techniques, it is possible to infer the position of the MS. For instance, as we can see in Figure 4.8, an extrapolation of the shapes of the silhouettes of landmarks can give some information about the location of the MS.

The video analysis technique, although it is conceptually simple to understand as it is the technique closest to the sense of orientation of a human being, may be extremely complex to implement due to the huge amount of computation and the complex algorithms that are involved in the procedure. However, this technique has the strong advantage of being still operative even if the MSs are in idle mode, because no emission of signals is required. For this reason, this technique is widely used in the context of robotics.

This technique encompasses two approaches: the use of a static scene, where only one image is analyzed at a time, and the use of a differential scene, where the analysis is done on the differences observed between sequential images. If the static scene is used, it is necessary to have a predefined database that maps the observed features in the image to the positions of the MSs. On the other hand, if the differential scene is used, the differences in the scene are related to the movement of an MS, allowing the MS to compute its location if the positions of the features in the image are known.

Figure 4.8 Video analysis technique.

4.4.3.3 Collaborative approach for database maintenance

Maintaining the database with information about wireless signals is not a simple task. One of the few automatic solutions relies on a large-scale collaborative approach. In a collaborative fashion, MSs which know their own locations report from time to time their locations and measurements obtained from the communication channel. This information, after processing, is then saved in the database as a fingerprint. Then, when a new user passes through the same location and requests his own location, this user is subsequently asked to report measurements obtained in its location. The system, by matching these measurements with known fingerprints, is able to return to the user the position stored in the database, if this position has previously been reported by another user. This technique is, for example, the approach used by My Location, which is a service provided by Google Maps for Mobile (Google 2007).

4.4.4 Dead reckoning

Dead reckoning (DR) is a technique that was widely used in the past in naval navigation (Bowditch 2002); it is currently used in advanced inertial systems (Farrell and Barth 1999) and, generally, is used in a range of different fields (Pantel and Wolf 2002). This technique is aimed at estimating the position of an MS or of any other type of target by predicting the position based on previously estimated positions or on FRPs. DR relies on pre-estimated or pre-acquired information regarding direction, speed, elapsed time and external factors such as frictional forces.

The DR concept is still widely used, integrated into more complex algorithms such as the Kalman filter described in Section 5.3.1. The prediction steps in movement-tracking algorithms often use a concept that resembles the DR method. Moreover, since terminal suppliers started to integrate inertial sensors into their smartphones, the range of applications of DR has increased. The combination of inertial sensors and wireless positioning solutions results in more complex systems, which certainly require some form of DR.

A serious problem of the technique when used without external information from FRPs is integration drift. This problem is related to a divergence of the predicted movement as a consequence of the accumulated errors that the DR returns in the various iterations of the algorithm. Thus, if no external information is available, the open-loop system that the DR represents is complex to handle.

In a hybrid scenario, this technique could be used, for instance, in combination with cell ID information. The DR algorithm would be responsible for predicting the movement of the MS while the cell ID information would function as FRPs in order to assist movement estimation.

4.4.5 Hybrid approaches

Hybrid approaches generally mean any combination of two or more techniques, not necessarily wireless positioning techniques. For instance, combining RSS-based triangulation methods with video camera motion tracking is considered a hybrid positioning technique. The major challenge is often the combination of data obtained from different techniques.

4.4.5.1 Hybrid angulation and lateration

The hybrid angulation and lateration (Venkatraman and Caffery 2004) technique is a straightforward technique that combines the two types of measurements by means of simply merging the formulas of each method.

Let us assume that n BSs, placed at $\{X_i = [x_i, y_i]^{\mathrm{T}}, i = 1, \ldots, n\}$, obtain TOA and AOA measurements from a communication channel with an MSand that each TOA measurement between BS_i and the MS is denoted by d_i and each AOA measurement between BS_i and the MS is denoted by θ_i. The hybrid angulation and lateration method can then be defined as:

$$X = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}B, \tag{4.31}$$

where

$$H = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_n - x_1 & y_n - y_1 \\ -\sin(\theta_1) & \cos(\theta_1) \\ -\sin(\theta_2) & \cos(\theta_2) \\ \vdots & \vdots \\ -\sin(\theta_n) & \cos(\theta_n) \end{bmatrix}, \tag{4.32}$$

and

$$B = \frac{1}{2} \begin{bmatrix} (d_1^2 - d_2^2) + (x_2^2 + y_2^2) - (x_1^2 + y_1^2) \\ (d_1^2 - d_3^2) + (x_3^2 + y_3^2) - (x_1^2 + y_1^2) \\ \vdots \\ (d_1^2 - d_n^2) + (x_n^2 + y_n^2) - (x_1^2 + y_1^2) \\ y_1 \cos(\theta_1) - x_1 \sin(\theta_1) \\ y_2 \cos(\theta_2) - x_2 \sin(\theta_2) \\ \vdots \\ y_n \cos(\theta_n) - x_n \sin(\theta_n) \end{bmatrix}. \tag{4.33}$$

4.4.5.2 Hybrid angulation and hyperbolic localization

Another example of hybrid localization is the combination of hyperbolic localization and angulation (Cong and Zhuang 2002). Let us assume that n BSs, placed at $\{X_i = [x_i, y_i]^{\mathrm{T}}, i = 1, \ldots, n\}$, obtain TDOA and AOA measurements from a communication channel with an MS and that each TDOA measurement between BS_i and BS_1 is denoted by t_i and can be related to a difference of distances d_i, and each AOA measurement between BS_i and the MS is denoted by θ_i. The hybrid angulation and hyperbolic-localization method can be defined as:

$$\mathsf{X} = [x, y, d_1]^{\mathrm{T}} = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} B, \tag{4.34}$$

where

$$H = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 & \mathsf{d}_2 \\ x_3 - x_1 & y_3 - y_1 & \mathsf{d}_3 \\ \vdots & \vdots & \vdots \\ x_n - x_1 & y_n - y_1 & \mathsf{d}_n \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ -\sin(\theta_2) & \cos(\theta_2) & 0 \\ \vdots & \vdots & \vdots \\ -\sin(\theta_n) & \cos(\theta_n) & 0 \end{bmatrix}, \tag{4.35}$$

and

$$B = \frac{1}{2}\begin{bmatrix} (x_2^2 + y_2^2) - (x_1^2 + y_1^2) - \mathsf{d}_2^2 \\ (x_3^2 + y_3^2) - (x_1^2 + y_1^2) - \mathsf{d}_3^2 \\ \vdots \\ (x_n^2 + y_n^2) - (x_1^2 + y_1^2) - \mathsf{d}_n^2 \\ y_1\cos(\theta_1) - x_1\sin(\theta_1) \\ y_2\cos(\theta_2) - x_2\sin(\theta_2) \\ \vdots \\ y_n\cos(\theta_n) - x_n\sin(\theta_n) \end{bmatrix}. \tag{4.36}$$

4.5 Error Sources in Positioning

The aim of this section is to briefly describe some of the most common types of errors that exist in the measurements of the communication channel and models for these errors. This section is based on the concepts introduced in Chapter 3.

4.5.1 Propagation

In wireless positioning, the most common problems and also the most complex problems to deal with are related to propagation effects, which often have a destructive effect on the measurements of the communication channel (Caffery and Stüber 1998). This section highlights some of these problems and describes some of their statistical properties.

4.5.1.1 Non-line-of-sight

The problem of NLOS paths is severe in range-based systems. The problem is due to obstructions between the transmitter and the receiver. These obstructions cause the signals to behave differently, as a consequence of the presence of several propagation media. The reason for this problem is related to differences in the properties characterizing the media that the signal propagates in. These properties include the velocity of propagation and the indexes of attenuation.

For instance, in RSS models such as those represented by Equation (4.2), the NLOS effect would be considered by including an additional attenuation l_{obstr}:

$$p = \alpha - 10\beta\log(d) + l_{\text{obstr}} + \eta, \quad \eta \sim \text{Norm}(0, \sigma). \tag{4.37}$$

This attenuation parameter l_{obstr} is strongly dependent on the propagation characteristics of the media that obstructs the line-of-sight communication. In more detail, this attenuation parameter can be defined as a loss given by a law similar to $l_{\text{obstr}} = \alpha_{\text{obstr}} - 10\beta_{\text{obstr}}\log(d_{\text{obstr}})$. Although the distance d_{obstr} of propagation through the obstruction maybe only of the order of few centimeters, for instance

in the case of a wall section, the propagation parameter in the wall material may have such a large propagation exponential β_{obstr} that the attenuation can be of the order of some tens of dB.

The NLOS effect is a difficult problem to solve in scenarios with several obstructions. A typical case is indoor scenarios, owing to the multiple obstructions among the rooms of a building, which cause effects that are difficult or even impossible to estimate in NLOS models. For this reason, NLOS effects strongly influence the positioning accuracy of range-based systems in indoor scenarios. Techniques for mitigating NLOS conditions in such cases are presented in Chapter 7.

4.5.1.2 Multipath fading

Multipath fading is an effect that corrupts signals at the receiver due to multiple reflections of copies of the same transmitted signal. These copies, because of their different propagation trajectories, arrive at the receiver with different attenuations, delays and phases compared with the direct signal. The copies of the signal are superimposed at the receiver, causing the properties of the signal to change.

Under NLOS conditions, it is common to model the envelope of the received signal by a Rayleigh distribution, i.e., the envelope tends with time to follow a Rayleigh distribution:

$$f_{\text{Rayleigh}}(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{4.38}$$

Given that the relation between the signal and the received power is quadratic (Goldsmith 2005), the Rayleigh distribution of the envelope results in an exponential distribution of the received power:

$$f_{\text{Rayleigh}}(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right). \tag{4.39}$$

According to this distribution of the received power, the model for the RSS in Equation (4.2) can be written as:

$$p = \alpha - 10\beta \log(d) + \eta, \quad \eta \sim \text{Exp}(\bar{\gamma}). \tag{4.40}$$

Under LOS conditions, the envelope of the received signal follows a Ricean distribution:

$$f_{\text{Rice}}(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + a^2}{2\sigma^2}\right) I_0\left(\frac{ar}{\sigma^2}\right), \tag{4.41}$$

$$\kappa = \frac{a^2}{2\sigma^2}, \tag{4.42}$$

where $I_0()$ is the modified Bessel function of the first kind of order 0, and the parameter a depends on the magnitude of the LOS ray. When $a = 0$, the distribution reduces to a Rayleigh distribution. In this case the Ricean distribution of the envelope

results in a complex distribution of the received power, which can be defined by:

$$f_{\text{Rice}}(\gamma) = \frac{2(1+\kappa)}{\bar{\gamma}} \exp(\kappa) \exp\left(-\frac{(1+\kappa)\gamma}{\bar{\gamma}}\right) I_0\left(\sqrt{\frac{4\kappa(1+\kappa)\gamma}{\bar{\gamma}}}\right). \tag{4.43}$$

4.5.1.3 Shadowing

Shadowing is an effect that is present when LOS conditions do not exist, but a nonobstructed ray is still possible by an indirect path. The signal needs to be reflected or refracted in order to travel along a nonobstructed path. This results in a longer distance traveled by the signal and a change in the angle from which the receiver listens to the signal (see Section 4.3.5). The measurements mostly affected by this effect are AOA measurements, although range-based measurements are also affected. Assuming that the target is static, an additional bias due to the diffraction of the signal is added to the AOA model represented by Equation (4.8):

$$\theta = \text{atan}\left(\frac{y_{\text{tx}} - y}{x_{\text{tx}} - x}\right) + \theta_{\text{bias}}, \tag{4.44}$$

where θ_{bias} is the bias introduced by the shadowing factor. When the device moves, the shadowing effect shows a probabilistic behavior. This behavior, however, is only visible for long movement distances, i.e., distances long enough that the major obstructions in the scenario (e.g., buildings, walls, vehicles) change. In this situation, the shadowing effect shows a log-normal distribution (Liberti and Rappaport 1992):

$$f_{\text{Shadowing}}(\gamma) = \frac{1}{\gamma\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(\gamma) - \mu)^2}{2\sigma^2}\right), \tag{4.45}$$

where μ and σ are, respectively, the mean and standard deviation of the logarithm of the variable γ. Since Equation (4.45) assumes γ to be in linear units, the corresponding equation in units of dB is given by a normal distribution:

$$f_{\text{Shadowing}}(\gamma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\gamma - \mu)^2}{2\sigma^2}\right). \tag{4.46}$$

4.5.1.4 Body shadowing

Body shadowing is also a severe problem for wireless communication technologies, in particular for wireless positioning (Obayashi and Zander 1998). This effect is due to the presence of the human body in close proximity to the MS. This presence introduces additional attenuations, commonly complex to model and characterized by large variability with time. Owing to this complexity, the destructive effect of body shadowing is commonly counteracted by increasing the transmission power. In fact, the impact of this effect in wireless positioning is currently a field of research under exploration.

4.5.1.5 Interference

Interference is another effect in wireless communications which has a complex statistical treatment, mainly because it strongly depends on the systems active in the vicinity of the MSs and the BSs. Also, the same technology is often a source of interference in the same system. Several aspects of the communication technology in use influence the communication channel, such as the transmitted power, MAC layer protocols and traffic, among other aspects. Besides the technology itself, the positions of all the transmitters and receivers also have an influence the interference. Owing to the complexity of the problem, interference is commonly studied in very specific scenarios, with strong assumptions, and it is typically treated in terms of average values.

4.5.1.6 The ionosphere

Ionospheric and tropospheric effects are strongly present in satellite positioning systems (Grewal et al. 2007). These effects are due to gases ionized by solar radiation. This ionization, besides producing ions, also produces clouds of free electrons, which impact on the propagation of the signals. Due to the periodic variation of the solar radiation intensity with the daily cycles of the Earth, the effects on the propagation also vary on a periodic basis, assuming that additional variations in solar activity are ignored. One aspect worth mentioning is that when a satellite is on the horizon, the signal travels a longer distance than in the case when the satellite is at the zenith. For this reason, robust ionospheric models must take into account the positions of the satellites.

The Klobuchar model, introduced by Klobuchar (1976) and analyzed in detail by Feess and Stephens (1987), is given by:

$$t_g = t_{\text{off}} + \mathsf{a}\left[1 - \frac{\omega^2}{2} + \frac{\omega^4}{24}\right], \quad \text{for } |\omega| \leq \frac{\pi}{2}, \tag{4.47}$$

where

$$\omega = \frac{2\pi\,(t_{\text{earth}} - t_{\text{phase}})}{\tau}, \tag{4.48}$$

with ω in radians. In Equation (4.47), the term t_{off} is a constant offset, generally assumed to be 5 ns; t_{phase} is a phase equal to 50 400 s (14 h); a is the amplitude and τ is the period of the ionospheric time delay; and t_{earth} the local time of the point on the Earth beneath the intersection of the signal with a mean ionospheric height of 350 km. The terms a and τ can be modeled by:

$$\mathsf{a} = \sum_{k=0}^{3} a_k \phi_m^k, \quad \tau = \sum_{k=0}^{3} b_k \phi_m^k, \tag{4.49}$$

where a_k and b_k are a set of coefficients (out of 370 sets) sent by the satellite to the user. The terms ϕ_m are the geomagnetic latitude of the ionospheric subpoint and can

be modeled by:

$$\phi_m = \phi_I + 11.6^\circ \cos(\lambda_I - 291^\circ), \tag{4.50}$$

where

$$\begin{aligned} \phi_I &= \phi_{\text{user}} + \theta_{\text{ea}} \cos(\theta_{\text{az}}), \\ \lambda_I &= \lambda_{\text{user}} + \theta_{\text{ea}} \frac{\cos(\theta_{\text{az}})}{\cos(\phi_I)}, \end{aligned} \tag{4.51}$$

and

$$\theta_{\text{ea}} = \frac{445}{l + 20^{-4}},$$

where $(\phi_{\text{user}}, \lambda_{\text{user}})$ are the geodesic latitude and longitude of the user, and (ϕ_I, λ_I) are the equivalent measures of the ionospheric subpoint. The terms θ_{ea}, θ_{az} and l are, respectively, the Earth angle, the azimuth and the elevation of the satellite with respect to the user.

4.5.2 Geometry

Another type of error widely known in the literature concerns the geometric distribution of the static FRPs. The problem is related to the relative positions between these FRPs and the target itself. This problem has been analyzed in the context of satellite positioning (Grewal et al. 2007). With the advent of positioning in cellular networks, the problem now also exists in this context.

Let us assume the position of the target to be X and the position of the n FRPs to be X_i, with $i = 1, \ldots, n$:

$$X = \begin{bmatrix} x & y & z \end{bmatrix}^{\text{T}}, \quad X_i = \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^{\text{T}}. \tag{4.52}$$

Since the clocks of the satellites and the receiver commonly have a constant offset, this offset c_{off} needs also to be calculated. Thus, the unknown position X of the receiver is extended with the clock offset c_{off}:

$$Y = \begin{bmatrix} X^{\text{T}} & c_{\text{off}} \end{bmatrix}^{\text{T}}. \tag{4.53}$$

It is possible to define the Euclidean distance between each FRP and the target as:

$$d_i(Y) \equiv d_i(X, c) = \sqrt{(X - X_i)^{\text{T}}(X - X_i)} + c_{\text{off}}, \tag{4.54}$$

$$D(Y) = \begin{bmatrix} d_1(Y) & \cdots & d_n(Y) \end{bmatrix}^{\text{T}}. \tag{4.55}$$

We take into account the fact that measurements are never noiseless, i.e., the measured ranges $\hat{D}(Y)$ are corrupted by a noise component η:

$$\hat{D}(Y) = D(Y) + \eta, \quad \eta = \begin{bmatrix} \eta_1 & \cdots & \eta_n \end{bmatrix}^{\text{T}}. \tag{4.56}$$

Expanding the term $D(Y)$ of Equation (4.56) as a Taylor series, we obtain:

$$\hat{D}(Y) = D(\hat{Y}) + \left.\frac{\partial D}{\partial Y}\right|_{\hat{Y}} (Y - \hat{Y}) + \eta, \quad \left.\frac{\partial D}{\partial Y}\right|_{\hat{Y}} = \left[\left.\frac{\partial D}{\partial X}\right|_{\hat{X}} \middle| \mathbf{1}\right], \tag{4.57}$$

where $\partial D/\partial Y|_{\hat{Y}}$ is the Jacobian matrix of $D(Y)$ for the values $\hat{Y}$, and $\mathbf{1}$ is the $n \times 1$ column vector of ones. By defining

$$\tilde{D} = \hat{D}(Y) - D(\hat{Y}), \quad \tilde{Y} = Y - \hat{Y}, \tag{4.58}$$

we have:

$$\tilde{D} = \left.\frac{\partial D}{\partial Y}\right|_{\hat{Y}} \tilde{Y} + \eta. \tag{4.59}$$

From the least-squares algorithm (presented in detail in Section 5.2), the value for $\tilde{Y}$ that minimizes the error η is given by:

$$\tilde{Y} = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}\tilde{D}, \quad H \equiv \left.\frac{\partial D}{\partial Y}\right|_{\hat{Y}}. \tag{4.60}$$

In order to understand the problem related to the geometry of the FRPs, let us assume that all the FRPs are concentrated in the same area with distances between them at least one order of magnitude smaller than the distance to the target. In this case, the entries in the Jacobian matrix H are of similar magnitude and, consequently, is close to singular. As a consequence, $\tilde{Y}$ increases. On the other hand, if the target is in the middle of the FRPs, the entries in the Jacobian H will have different magnitudes, and the result is that H is far from singularity; the result is then a lower value for $\tilde{Y}$.

4.5.3 Equipment and technology

Another type of problem that influences wireless positioning accuracy concerns the equipment itself and the technologies with which positioning is performed. The impact of these errors can be seen in several different ways, such as in variability in the amplitude of errors, biases, non-stationary measurement processes and variations in the distribution of errors. The most common type of error that influences the communication is thermal noise, which causes fluctuations in the measurements used later in position estimation.

The technology itself can also introduce errors in the measurements. For instance, MAC protocols with randomized processes, such as collision avoidance or listen-before-talk techniques, commonly used to circumvent interference problems, can also introduce additional delays in the measurements used for wireless positioning. The IEEE 802.11 technology is one case. These delays are related to the timestamps of the measurements rather than to the actual time measurements used for positioning. Furthermore, the resolution of the measurements permitted in the technology may also be a source of errors. See, for instance, the Bluetooth case (see Section 8.4.1.2),

where by specification it is only possible to get measurements with a granularity of 1 dB. Concerning time measurements, clock synchronization errors are a major problem, introducing reading errors into TOA and TDOA measurements. Cellular communication systems such as GSM and UMTS are subject to clock errors.

Apart from the communication technology, even the network management strategy can introduce errors or misinformation into positioning applications. A typical example occurs in cellular networks where the operator often treat the locations of the BSs as confidential, meaning that the estimation of those locations commonly has to be performed using opportunistic positioning solutions. Furthermore, the BS cell IDs provided by the operators to the MSs are often incomplete, which may result in confusion between measurements obtained from different BSs *that share the same information about the known part of the incomplete information*.

The range of sources of errors is indeed large and sometimes even unexpected by positioning-system designers. Besides the examples given, there are many others, such as governmental requirements to degrade the accuracy of GPS data and design decisions of MS providers that do not provide APIs for obtaining RSS measurements, or restrict those APIs.

4.6 Metrics of Location Accuracy

This section is aimed at describing some of the metrics used to estimate accuracy levels for positioning applications. There are several ways of characterizing accuracy in the literature; however, in this section we state some standard metrics used in positioning applications.

4.6.1 Circular error probability

The circular error probability (CEP) is a metric that was first defined in the context of military ballistics. It is a simple measure of precision that defines a circle within which the probability of there being a correct estimate of the position is 50%:

$$d : \mathrm{P}(\|X, \hat{X}\| < d) = 0.5. \tag{4.61}$$

4.6.2 Dilution of precision

DOP is a concept that is tightly related to GPS positioning, but extendable to any range-based positioning solution. DOP gives information about the impact that the configuration of the satellites or FRPs has on positioning accuracy. In order to understand the concept of DOP (Grewal et al. 2007), it is necessary to start by calculating the error covariance of $\tilde{Y}$ in Equation (4.60):

$$\mathrm{E}[\tilde{Y}\tilde{Y}^{\mathrm{T}}] = \mathrm{E}[(H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}\tilde{D}[(H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}\tilde{D}]^{\mathrm{T}}], \tag{4.62}$$

$$\mathrm{E}[\tilde{Y}\tilde{Y}^{\mathrm{T}}] = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}\mathrm{E}[\tilde{D}\tilde{D}^{\mathrm{T}}]H(H^{\mathrm{T}}H)^{-1}. \tag{4.63}$$

Let us assume that there is no correlation between the measurements obtained from each of the n FRPs and that such measurements are corrupted by a noise component with variance σ^2:

$$\mathrm{E}[\tilde{Y}\tilde{Y}^{\mathrm{T}}] = \sigma^2 (H^{\mathrm{T}}H)^{-1} \underbrace{H^{\mathrm{T}}H(H^{\mathrm{T}}H)^{-1}}_{I}, \tag{4.64}$$

$$\mathrm{E}[\tilde{Y}\tilde{Y}^{\mathrm{T}}] = \sigma^2 (H^{\mathrm{T}}H)^{-1}. \tag{4.65}$$

As Equation (4.65) shows, the term $(H^{\mathrm{T}}H)^{-1}$ establishes a direct relationship between the error covariance in the measurements $d_i(Y)$ and the error covariance in the position coordinates Y. For this reason, the term $(H^{\mathrm{T}}H)^{-1}$ is responsible for the dilution of precision in the positioning of the receiver. Depending on the entries in the term $(H^{\mathrm{T}}H)^{-1}$, it is possible to determine several different types of DOP:

$$\epsilon_{\mathrm{GDOP}} = \sqrt{a_{11} + a_{22} + a_{33} + a_{44}}, \tag{4.66}$$

$$\epsilon_{\mathrm{PDOP}} = \sqrt{a_{11} + a_{22} + a_{33}}, \tag{4.67}$$

$$\epsilon_{\mathrm{HDOP}} = \sqrt{a_{11} + a_{22}}, \tag{4.68}$$

$$\epsilon_{\mathrm{VDOP}} = \sqrt{a_{33}}, \tag{4.69}$$

and

$$\epsilon_{\mathrm{TDOP}} = \sqrt{a_{44}}, \tag{4.70}$$

where the scalars a are the entries in the term $(H^{\mathrm{T}}H)^{-1}$. From Equations (4.66)–(4.70), ϵ_{GDOP}, ϵ_{PDOP}, ϵ_{HDOP}, ϵ_{VDOP} and ϵ_{TDOP} are the geometric, position, horizontal, vertical and time DOP metrics, respectively. While, for instance, ϵ_{PDOP} accounts for 3D position errors, ϵ_{TDOP} accounts only for time errors.

4.6.3 Cramér–Rao lower bound (CRLB)

The CRLB establishes a lower limit for the variance that an unbiased estimator of a parameter of a distribution can attain (Deffenbaugh et al. 1996). Let θ be an unbiased, unknown column vector of parameters of the distribution $p(r, \theta)$ of a random variable r. If an estimator of θ is denoted by $\hat{\theta}$, the CRLB is given by:

$$\mathrm{var}(\hat{\theta}) \geq I(\theta)^{-1}, \tag{4.71}$$

where the inequality $A \geq B$ is equivalent to saying that the matrix $A - B$ is positive semi-definite. The term $I(\theta)$ represents the Fisher information matrix (FIM), which is given by:

$$I(\theta) = -E\left[\frac{\partial^2 \log\, p(r, \theta)}{\partial \theta^2}\right]. \tag{4.72}$$

Since by definition the mean squared error (MSE) is given by the variance of an unbiased estimator, Equation (4.71) corresponds also to establishing a limit on

the MSE. In a more general case where instead of estimating θ one is interested in estimating a function $g(\theta)$, assuming that this estimator has a bias $b(\theta)$, the MSE is given by:

$$\text{MSE} \leq \frac{\partial(g+b)(\theta)}{\partial\theta} I(\theta)^{-1} \left(\frac{\partial(g+b)(\theta)}{\partial\theta} \right)^{\text{T}} + b^2(\theta). \tag{4.73}$$

4.7 Conclusions

This chapter has reviewed the fundamentals of wireless positioning. First, positioning techniques were classified according to various criteria: the topology of the positioning system, the physical coverage range and the level of integration in the communication technology. The types of measurements are typically: RSS, TOA, TDOA and AOA, while the typical positioning techniques rely on cell identification, triangulation or fingerprinting strategies. Several error sources and models were illustrated mainly for propagation-related effects. The chapter closed by defining some of the most common accuracy metrics. Practically, the aim of this chapter was to give the basics needed to prepare the reader for the forthcoming chapters, which deal with more advanced techniques for wireless positioning.

5

Data Fusion and Filtering Techniques

5.1 Introduction

As has been widely presented in previous chapters, wireless channel measurements for position estimation are unavoidably corrupted by noise. As described in Chapter 3, the various types of noise sources span a wide range of effects, commonly causing the channel measurements to present statistical behaviors that are difficult to model and complex to circumvent. Such complexity results in observable impairments in the positioning mechanisms, which impact on the accuracy of the system. There are several approaches to wireless positioning that are aimed at handling the statistical behavior of the measurement noise. In this chapter, two major approaches are presented: the least-squares (LS) approach and the Bayesian framework.

The LS approach is an algorithm that estimates the position of an MS by minimizing the squared error between the actual measurements observed in the wireless channel and the expected measurements resulting from the estimated position. In order to execute this algorithm, it is necessary to model the relation between the position and the actual measurements. We present both the linear and the nonlinear case of the LS approach. As a particular case of the linear LS, the recursive version is presented.

In the Bayesian framework, the position is determined as an estimator that minimizes the mean square error between the actual measurements and the expected measurements. Three major implementations of the Bayesian framework are presented: the Kalman filter (KF), the particle filter (PF) and grid-based methods. The KF is a group of methods that assumes the measurements to be corrupted by white Gaussian noise, the PF is a Monte Carlo type of algorithm that does

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

not constrain the noise component of the Gaussian distribution, and the grid-based methods assume that the state space is discrete and finite.

In addition to these two major approaches to data fusion, alternative solutions are also analyzed. The first alternative solution relies on fingerprinting techniques. The fingerprint technique follows a prephase of calibration where measurements are obtained at several strategic positions. These measurements are then used in the operational mode as reference values to be correlated with future measurements. The second alternative relies on time series theory which by averaging pre-determined position estimators over time tries to minimize the effect of errors.

5.2 Least-squares Methods

LS methods are used for calculating the solutions of overdetermined systems. In statistical contexts, LS methods can be used to approximate a given set of data points by estimating the parameters of a predefined model. In the context of wireless positioning, these methods can be used to approximate a set of measurements from the wireless channel by estimating the position of a certain MS. The estimated position is obtained by minimizing the sum of the squared errors between a model that relates the position to wireless measurements and the actual data measured from the wireless channel.

Let us assume that the position coordinates $X = [x, y]^{\mathrm{T}}$ in a two-dimensional space correspond to the channel measurements $Z = [z_1, \ldots, z_n]^{\mathrm{T}}$ at the ith BS, where n is the number of BSs. Let $h_i(X)$ be a model that relates the two variables such that

$$z_i = h_i(X) + \eta_i. \tag{5.1}$$

The parameter η_i represents a noise component added to the model. The best fit according to the least-squares approach is when the cost function $\mathcal{J}(X)$ is minimum:

$$\mathcal{J}(X) = \sum_{i=1}^{n} \eta_i^2 = \sum_{i=1}^{n} [z_i - h_i(X)]^2 = [Z - \mathcal{H}(X)]^{\mathrm{T}} [Z - \mathcal{H}(X)], \tag{5.2}$$

where $\mathcal{H}(X) = [h_1(X), \ldots, h_n(X)]^{\mathrm{T}}$. The estimated position $\hat{X}$ is determined as the argument that minimizes Equation (5.2), i.e.,

$$\hat{X} = \arg \min_{X} \{\mathcal{J}(X)\}. \tag{5.3}$$

The solution to Equation (5.3) can be obtained by setting the gradient of $\mathcal{J}(X)$ to zero, where the partial derivatives of $\mathcal{J}(X)$ are taken with respect to X:

$$\nabla \mathcal{J}(X) = -2[Z - \mathcal{H}(X)]^{\mathrm{T}} \nabla \mathcal{H}(X) = 0. \tag{5.4}$$

Equation (5.4) defines the framework to be used in an LS method. Once the model $h_i(X)$ is known, it may be possible to manipulate Equation (5.4) further into a

simpler solution; this is the case when the model is linear. In other cases, numerical optimization may be necessary.

A more general approach to the LS method assumes that the various sources of data produce measurements corrupted by error components with different statistical properties with respect to the other sources. In this general case, the method is given the name of weighted least squares (WLS). Thus, assuming that the measurement z_i is corrupted by white Gaussian noise with standard deviation σ_i, i.e., $\eta_i \sim \text{Norm}(0, \sigma_i)$ with $\eta = [\eta_1, \ldots, \eta_n]^{\text{T}}$ and $\text{E}[\eta\eta^{\text{T}}] = R$, the best fit in the least-squares sense is when the cost function $\mathcal{J}(X)$ is minimum:

$$\mathcal{J}(X) = \sum_{i=1}^{n} R_{i,i}^{-1}[z_i - h_i(X)]^2 = [Z - \mathcal{H}(X)]^{\text{T}} R^{-1}[Z - \mathcal{H}(X)]. \tag{5.5}$$

The minimum is then obtained from:

$$\nabla\mathcal{J}(X) = -2[Z - \mathcal{H}(X)]^{\text{T}} R^{-1} \nabla\mathcal{H}(X) = 0. \tag{5.6}$$

In Equation (5.6), it is assumed that R is a symmetric matrix. Commonly, R is in fact defined as a diagonal matrix, i.e., the measurements are assumed to be independent from each other.

5.2.1 Linear least squares

A special case of the LS methods is the linear least-squares (LLS) method. The peculiarity of this method is that the model $\mathcal{H}(X)$ in Equation (5.4) is assumed to be linear, i.e.,

$$\mathcal{H}(X) = HX. \tag{5.7}$$

This linearity of the model simplifies the mathematical formulation of the method, in the sense that it is possible to get closed-form solutions for the parameters that minimize the cost function of Equation (5.2), i.e., it is possible to obtain a closed-form solution for Equation (5.4). Thus, by using Equation (5.7) in Equation (5.4), we have:

$$\nabla\mathcal{J}(X) = -2[Z^{\text{T}} - X^{\text{T}} H^{\text{T}}] \nabla HX = 0. \tag{5.8}$$

Since the gradient of $\mathcal{H}(X)$ with respect to the parameters X equals H, Equation (5.8) results in:

$$\nabla\mathcal{J}(X) = -2[Z^{\text{T}} - X^{\text{T}} H^{\text{T}}] H = 0. \tag{5.9}$$

Then, after further manipulation:

$$X^{\text{T}} H^{\text{T}} H = Z^{\text{T}} H, \tag{5.10}$$

and, after transposing both sides of the equation, we obtain:

$$H^{\text{T}} H X = H^{\text{T}} Z. \tag{5.11}$$

It is now possible to get a closed-form solution for the LLS method:

$$X = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}Z. \tag{5.12}$$

As Equation (5.12) shows, the LLS method can be used easily as a minimization tool. In particular, for positioning applications, the LLS algorithm can be used to determine the position of an MS when there are measurements available from several BSs. This is the example illustrated in Sections 3.6 and 3.7.

For the weighted formulation of the LLS method, similar manipulations can be done. Owing to its straightforwardness, the calculations will not be presented here. The weighted LLS solution is given by:

$$X = (H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}R^{-1}Z. \tag{5.13}$$

An important property of the LLS method is the covariance matrix of the LS estimator. This covariance matrix represents a measurement of the certainty in the estimation of the position:

$$P = \mathrm{E}[\{\hat{X} - \mathrm{E}[\hat{X}]\}\{\hat{X} - \mathrm{E}[\hat{X}]\}^{\mathrm{T}}] \tag{5.14}$$

$$= \mathrm{E}[\{\hat{X} - X\}\{\hat{X} - X\}^{\mathrm{T}}] \tag{5.15}$$

$$= (H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}R^{-1}\mathrm{E}[\eta\eta^{\mathrm{T}}]R^{-1}H(H^{\mathrm{T}}R^{-1}H)^{-1} \tag{5.16}$$

$$= (H^{\mathrm{T}}R^{-1}H)^{-1}H^{\mathrm{T}}R^{-1}RR^{-1}H(H^{\mathrm{T}}R^{-1}H)^{-1}. \tag{5.17}$$

Equation (5.17) results in:

$$P = (H^{\mathrm{T}}R^{-1}H)^{-1}. \tag{5.18}$$

5.2.2 Recursive least squares

The recursive least-squares (RLS) algorithm is a recursive method for iteratively determining the LLS estimator as a function of time. The RLS algorithm typically updates the current estimator by running each iteration at every discrete time k when a newly obtained measurement is available. The following matrices are used to ease the manipulation:

$$Z_{0:k+1} = \begin{bmatrix} Z_{0:k} \\ Z_{k+1} \end{bmatrix}, \tag{5.19}$$

$$H_{0:k+1} = \begin{bmatrix} H_{0:k} \\ H_{k+1} \end{bmatrix} \tag{5.20}$$

and

$$R_{0:k+1} = \begin{bmatrix} R_{0:k} & 0 \\ 0 & R_{k+1} \end{bmatrix}. \tag{5.21}$$

Taking the covariance matrix as a starting point, it is possible to write:

$$P_{k+1}^{-1} = H_{0:k+1}^{\mathrm{T}}(R_{0:k+1})^{-1}H_{0:k+1} \tag{5.22}$$

$$= \begin{bmatrix} H_{0:k}^{\mathrm{T}} & H_{k+1}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} R_{0:k} & 0 \\ 0 & R_{k+1} \end{bmatrix}^{-1} \begin{bmatrix} H_{0:k} \\ H_{k+1} \end{bmatrix} \tag{5.23}$$

$$= H_{0:k}^{\mathrm{T}}R_{0:k}^{-1}H_{0:k} + H_{k+1}^{\mathrm{T}}R_{k+1}^{-1}H_{k+1} \tag{5.24}$$

$$= P_k^{-1} + H_{k+1}^{\mathrm{T}}R_{k+1}^{-1}H_{k+1}, \tag{5.25}$$

where the errors in the measurements are not correlated in time. Using the matrix inversion lemma, Equation (5.25) can be rewritten as:

$$P_{k+1} = [P_k^{-1} + H_{k+1}^{\mathrm{T}}R_{k+1}^{-1}H_{k+1}]^{-1} \tag{5.26}$$

$$= P_k - P_kH_{k+1}^{\mathrm{T}}[H_{k+1}P_kH_{k+1}^{\mathrm{T}} + R_{k+1}]^{-1}H_{k+1}P_k. \tag{5.27}$$

This results in the following equation for P_{k+1}:

$$P_{k+1} = [I - K_{k+1}H_{k+1}]P_k, \tag{5.28}$$

where K_{k+1} is the update gain, which is given by:

$$K_{k+1} = P_kH_{k+1}^{\mathrm{T}}[H_{k+1}P_kH_{k+1}^{\mathrm{T}} + R_{k+1}]^{-1}. \tag{5.29}$$

In Equation (5.29), the term $H_{k+1}P_kH_{k+1}^{\mathrm{T}} + R_{k+1}$ is the covariance of the residuals. As it is demonstrated in Bar-Shalom et al. (2001), the update gain K_{k+1} can also be written as $K_{k+1} = P_{k+1}H_{k+1}^{\mathrm{T}}R_{k+1}^{-1}$, which formulation is of clear interest in the manipulation of the update formulas for the position estimator X shown below.

Using Equations (5.18) and (5.19)–(5.21), the batch LS estimator of Equation (5.13) can be rewritten as:

$$\hat{X}_{k+1} = P_{k+1}H_{0:k+1}^{\mathrm{T}}(R_{0:k+1})^{-1}Z_{0:k+1} \tag{5.30}$$

$$= P_{k+1}\begin{bmatrix} H_{0:k}^{\mathrm{T}} & H_{k+1}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} R_{0:k} & 0 \\ 0 & R_{k+1} \end{bmatrix}^{-1} \begin{bmatrix} Z_{0:k} \\ Z_{k+1} \end{bmatrix} \tag{5.31}$$

$$= P_{k+1}H_{0:k}^{\mathrm{T}}(R_{0:k})^{-1}Z_{0:k} + P_{k+1}H_{k+1}^{\mathrm{T}}R_{k+1}^{-1}Z_{k+1} \tag{5.32}$$

$$= [I - K_{k+1}H_{k+1}]P_kH_{0:k}^{\mathrm{T}}(R_{0:k})^{-1}Z_{0:k} + K_{k+1}Z_{k+1} \tag{5.33}$$

$$= [I - K_{k+1}H_{k+1}]\hat{X}_k + K_{k+1}Z_{k+1}, \tag{5.34}$$

and, consequently,

$$\hat{X}_{k+1} = \hat{X}_k + K_{k+1}[Z_{k+1} - H_{k+1}\hat{X}_k]. \tag{5.35}$$

The RLS algorithm is defined by Equations (5.35), (5.28) and (5.29).

5.2.3 Weighted nonlinear least squares

The previous sections have shown examples of LS methods for the particular case where the model in Equation (5.4) is linear. While for linear models the LS method can be mathematically manipulated in order to obtain a closed-form solution, in the generic case of nonlinear models, the mathematical formulations are more complex. The typical solution for the nonlinear case is to solve the LS problem using a numerical method for optimization. The problem can be seen as a numerical optimization of:

$$\hat{X} = \arg\min_X \{\mathcal{J}(X)\} = \arg\min_X \left\{ \sum_{i=1}^{n} \gamma_i [z_i - h_i(X)]^2 \right\}, \tag{5.36}$$

where γ_i is a weight, commonly defined as the inverse of the variance.

For instance, the cost function to be minimized when TOA and AOA measurements are available is[1]:

$$\mathcal{J}(X) = \sum_{i=1}^{n} \gamma_t^{[i]} (\eta_t^{[i]})^2 + \gamma_\theta^{[i]} (\eta_\theta^{[i]})^2, \tag{5.37}$$

where $\eta_t^{[i]}$ and $\eta_\theta^{[i]}$ represent the error components of the measurements of TOA and AOA, respectively, and $\gamma_t^{[i]}$ and $\gamma_\theta^{[i]}$ are weights appropriately selected to represent the reliability of the TOA and AOA measurements, respectively. When TDOA and AOA measurements are available, the cost function is:

$$\mathcal{J}(X) = \sum_{i=1}^{n} \gamma_{\mathrm{t}}^{[i]} (\eta_{\mathrm{t}}^{[i]})^2 + \gamma_\theta^{[i]} (\eta_\theta^{[i]})^2, \tag{5.38}$$

where $\eta_{\mathrm{t}}^{[i]}$ is the error component for a measurement of TDOA and $\gamma_{\mathrm{t}}^{[i]}$ is the corresponding reliability weight.

In Equations (5.37) and (5.38), the error functions $\eta_t^{[i]}$, $\eta_{\mathrm{t}}^{[i]}$ and $\eta_\theta^{[i]}$ are respectively defined as:

$$\eta_t^{[i]} = t^{[i]} - \sqrt{[X - X^{[i]}]^{\mathrm{T}}[X - X^{[i]}]}/c, \tag{5.39}$$

$$\eta_{\mathrm{t}}^{[i]} = \mathrm{t}^{[i]} - \left(\sqrt{[X - X^{[i]}]^{\mathrm{T}}[X - X^{[i]}]} - \sqrt{[X - X^{[1]}]^{\mathrm{T}}[X - X^{[1]}]}\right)/c \tag{5.40}$$

and

$$\eta_\theta^{[i]} = \theta^{[i]} - \arctan(X - X^{[i]}), \tag{5.41}$$

where c is the speed of light, $X^{[i]}$ is the location coordinates of the ith BS, and $t^{[i]}$, $\mathrm{t}^{[i]}$ and $\theta^{[i]}$ are the TOA, TDOA and AOA measurements respectively, at the ith BS.[2] The

[1] The unconstrained minimization procedure employed to search for the minimum of the objective function is the trust region method defined in Coleman and Li (1996).

[2] If we have available more than a single TOA, TDOA or AOA measurement, $t^{[i]}$, $\mathrm{t}^{[i]}$ and $\theta^{[i]}$ are the median values calculated over each set of measurements.

weights $\gamma_t^{[i]}$, $\gamma_{\mathrm{t}}^{[i]}$ and $\gamma_\theta^{[i]}$ in Equations (5.37) and (5.38) are appropriately selected in order to reflect the reliability of the TOA, TDOA and AOA measurements:

$$\gamma_t^{[i]} = (\sigma_t^{[i]})^{-2}, \tag{5.42}$$

$$\gamma_{\mathrm{t}}^{[i]} = (\sigma_{\mathrm{t}}^{[i]})^{-2} \tag{5.43}$$

and

$$\gamma_\theta^{[i]} = (\sigma_\theta^{[i]})^{-2}, \tag{5.44}$$

where $\sigma_t^{[i]}$, $\sigma_{\mathrm{t}}^{[i]}$ and $\sigma_\theta^{[i]}$ are directly derived from the available measurements and represent the standard deviation of the error in the measurements.

5.2.3.1 Example of application

Let us consider an MS in range of a geographically distributed number of cellular BSs ($BS_1, \ldots, BS_n$). We assume that for each available BS, the MS performs m observations of either TOA and TDOA and that the home BS, BS_1, performs m observations of AOA on the communications link BS_1–MS. After the estimation process, at the location server we have the following set of available data: (i) (**TOA**, **AOA**) or (ii) (**TDOA**, **AOA**), with entries defined by:

$$\mathbf{TOA} = \begin{bmatrix} \mathbf{TOA}_1 \\ \vdots \\ \mathbf{TOA}_n \end{bmatrix} = \begin{bmatrix} TOA_{1,1} \cdots TOA_{1,m} \\ \vdots \\ TOA_{n,1} \cdots TOA_{n,m} \end{bmatrix}, \tag{5.45}$$

$$\mathbf{TDOA} = \begin{bmatrix} \mathbf{TDOA}_1 \\ \vdots \\ \mathbf{TDOA}_{n-1} \end{bmatrix} = \begin{bmatrix} TDOA_{2,1} \cdots TDOA_{2,m} \\ \vdots \\ TDOA_{n,1} \cdots TDOA_{n,m} \end{bmatrix} \tag{5.46}$$

and

$$\mathbf{AOA} = [\mathbf{AOA}_1] = [AOA_{1,1} \cdots AOA_{1,m}]. \tag{5.47}$$

Note that since TDOA measurements need two BSs and those measurements are taken with respect to BS_1, the number of actual observations is $n - 1$. For a better understanding, see Section 4.4.2.2. Let us write this problem in a WNLLS form.

STEP 0: In order to efficiently describe our problem, we need to develop a mathematical formalism. Therefore, we start by introducing the following definitions

(Frattasi and Monti 2007b):[3]

$$X^{[j]} = \begin{bmatrix} x^{[j]} \\ y^{[j]} \end{bmatrix}, \quad j = 1, \ldots, n, \tag{5.48}$$

$$\hat{X} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}, \tag{5.49}$$

where $X^{[j]}$ represents the location coordinates of BS_j and $\hat{X}$ represents the estimated location coordinates of the MS, both in a Cartesian coordinate system:

$$\hat{d}^{[j]} = (t^{[j]} - t_{\mathrm{tx}})c, \quad j = 1, \ldots, n, \tag{5.50}$$

$$\hat{\mathsf{d}}^{[j]} = (t^{[j]} - t_{\mathrm{tx}})c - (t^{[1]} - t_{\mathrm{tx}})c = (t^{[j]} - t^{[1]})c = \mathsf{t}^{[j]}c, \quad j \neq 1, \tag{5.51}$$

where $\hat{d}^{[j]}$ is the estimated distance between the MS and BS_j derived from $t^{[j]}$, the TOA of BS_j's signal at the MS; $\hat{\mathsf{d}}^{[j]}$ is the estimated distance difference between the MS and the pair (BS_1, BS_j) derived from $\mathsf{t}^{[j]}$, the TDOA of BS_1 and BS_j's signals at the MS; $c = 3 \times 10^8$ m/s is the speed of light; and t_{tx} is the time instant at which BS_1 and BS_j begin to transmit. In addition,

$$\hat{d}_k^{[j]} = \sqrt{\{\hat{X}_{k-1} - X^{[j]}\}^{\mathrm{T}}\{\hat{X}_{k-1} - X^{[j]}\}}, \quad j = 1, \ldots, n, \tag{5.52}$$

$$\hat{\mathsf{d}}_k^{[j]} = \sqrt{\{\hat{X}_{k-1} - X^{[j]}\}^{\mathrm{T}}\{\hat{X}_{k-1} - X^{[j]}\}} - \sqrt{\{\hat{X}_{k-1} - X^{[1]}\}^{\mathrm{T}}\{\hat{X}_{k-1} - X^{[1]}\}}, \; j \neq 1, \tag{5.53}$$

$$\hat{\theta}_k^{[1]} = \arctan\left(\frac{\hat{y}_{k-1} - y^{[1]}}{\hat{x}_{k-1} - x^{[1]}}\right), \tag{5.54}$$

where $\hat{d}_k^{[j]}$ and $\hat{\mathsf{d}}_k^{[j]}$ have meanings similar to those in Equations (5.50) and (5.51), but with the difference that here, instead of being calculated directly from the TOA or TDOA measurements, they are derived from $\hat{X}_{k-1} = [\hat{x}_{k-1}, \hat{y}_{k-1}]^{\mathrm{T}}$, the estimated location coordinates of the MS at iteration $k-1$ of the minimization routine. The same applies to $\hat{\theta}_k^{[1]}$, the estimated angle between the MS and BS_1.

STEP 1: The WNLLS method requires an initial guess $\hat{X}_0 = [\hat{x}_0, \hat{y}_0]^{\mathrm{T}}$ at the beginning of its minimization routine.

STEP 2: Unlike the KF method (see Section 5.3.1), which picks up in each iteration a new set of measurements (e.g., ($\mathbf{TOA}_{1,1}$, $\mathbf{AOA}_{1,1}$), ($\mathbf{TOA}_{1,2}$, $\mathbf{AOA}_{1,2}$), etc.), the WNLLS method uses only one set of measurements at the beginning of

[3] For simplicity of presentation, only the x and y coordinates are considered in the derivations given here (this corresponds to the case where the MS and BS_j are located on a relatively flat plane). Although the z coordinate is ignored, the present example can easily be extended to the three-dimensional case.

its minimization routine. As a consequence, in order for the WNLLS method to be able to exploit all of the available data wisely, we apply the median operator, i.e., $\widetilde{(\bullet)}$, to each of the rows in Equations (5.45)–(5.47) and, by using the formalism introduced previously, we obtain[4]:

$$\mathbf{TOA} = \begin{bmatrix} \widetilde{\mathbf{TOA}}_1 \\ \vdots \\ \widetilde{\mathbf{TOA}}_n \end{bmatrix} = \begin{bmatrix} t^{[1]} \\ \vdots \\ t^{[n]} \end{bmatrix}, \tag{5.55}$$

$$\mathbf{TDOA} = \begin{bmatrix} \widetilde{\mathbf{TDOA}}_1 \\ \vdots \\ \widetilde{\mathbf{TDOA}}_{n-1} \end{bmatrix} = \begin{bmatrix} \mathsf{t}^{[2]} \\ \vdots \\ \mathsf{t}^{[n]} \end{bmatrix} \tag{5.56}$$

and

$$\mathbf{AOA} = [\widetilde{\mathbf{AOA}}_1] = [\theta^{[1]}]. \tag{5.57}$$

STEP 3: Finally, we can write the objective function $\mathcal{J}(X)$ in Equation (5.36) for each of the two envisioned cases, i.e., (**TOA**, **AOA**) and (**TDOA**, **AOA**), as (Frattasi 2007):

$$\mathcal{J}(X) = \sum_{j=1}^{n} \gamma_t^{[j]} \{f_t(\hat{X})\}^2 + \gamma_\theta^{[1]} \{f_\theta(\hat{X})\}^2 \tag{5.58}$$

and

$$\mathcal{J}(X) = \sum_{j=1}^{n} \gamma_{\mathsf{t}}^{[j]} \{f_{\mathsf{t}}(\hat{X})\}^2 + \gamma_\theta^{[1]} \{f_\theta(\hat{X})\}^2, \tag{5.59}$$

where the functions $f(\bullet)$ are defined as:

$$f_t(\hat{X}) = \hat{d}^{[j]} - \hat{d}_k^{[j]}, \quad j = 1, \ldots, n, \tag{5.60}$$

$$f_{\mathsf{t}}(\hat{X}) = (\hat{d}^{[j]} - \hat{d}^{[1]}) - (\hat{d}_k^{[j]} - \hat{d}_k^{[1]}) = \hat{\mathsf{d}}^{[j]} - \hat{\mathsf{d}}_k^{[j]}, \quad j = 1, \ldots, n, \tag{5.61}$$

and

$$f_\theta(\hat{X}) = \theta^{[1]} - \hat{\theta}_k^{[1]}. \tag{5.62}$$

The weights γ_t, γ_{t} and γ_θ should be appropriately selected in order to reflect the reliability of the available TOA, TDOA and AOA measurements. In particular, they can be chosen as (Frattasi 2007):

$$\gamma_t^{[j]} = \frac{1}{\sigma^2_{\mathbf{TOA}_j}}, \quad j = 1, \ldots, n, \tag{5.63}$$

$$\gamma_{\mathsf{t}}^{[j]} = \frac{1}{\sigma^2_{\mathbf{TDOA}_j}}, \quad j = 1, \ldots, n, \tag{5.64}$$

[4]The *median* operator is preferred to the *mean* operator, owing to its higher robustness against biased measurements.

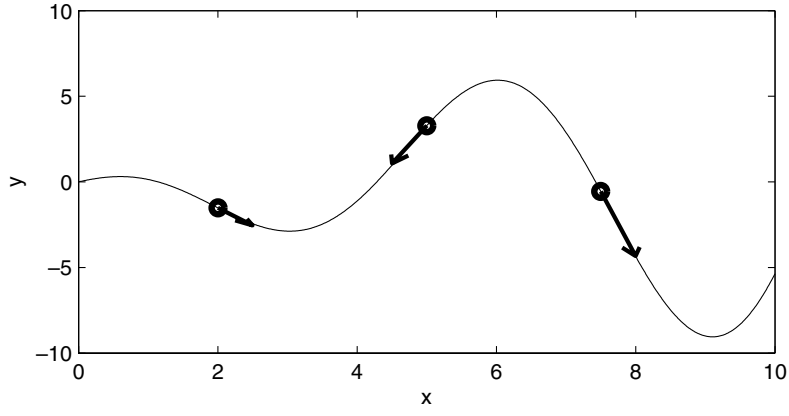


Figure 5.1 Local-minimum problem (the function has been chosen for illustrative purposes only).

and

$$\gamma_{\theta}^{[1]} = \frac{1}{\sigma^2_{\mathbf{AOA}_1}}, \tag{5.65}$$

where $\sigma^2_{\mathbf{TOA}_j}$, $\sigma^2_{\mathbf{TDOA}_j}$ and $\sigma^2_{\mathbf{AOA}_1}$ are variances directly derived from the available sets of data.

5.2.4 The absolute/local-minimum problem

Optimization algorithms typically operate in a given interval and iteratively evaluate the values of the subject function in consecutive smaller subintervals until a certain criteria of decision is met. The requirement is that the subject function shall assure that inside that interval the problem is convex, otherwise the optimization algorithm does not assure that the absolute minimum is determined. For instance, given the interval in the example depicted in Figure 5.1, the optimization algorithm could evolve in the direction pointed by the arrows. As convexity is not assured within that interval, the local minimum around $x = 3$ could be returned by the optimization algorithm instead of the absolute minimum around $x = 9$. As it is complex or even impossible to circumvent this problem and assure convergence to an absolute minimum, system designed shall take this fact into consideration.

5.3 Bayesian Filtering

Bayesian filtering techniques are techniques that estimate the state of a dynamic system based on measurements from noisy sensors (Arulampalam et al. 2002; Grewal and Andrews 2001). These techniques are widely used in several contexts, such as business and economics, signals and systems, image processing and mobile tracking.

In the specific case of location information estimation in scenarios with static or moving devices, Bayesian filtering techniques have been successfully applied to solve the problem of noisy measurements in localization applications (Fox et al. 2003).

The Bayesian filter is an abstract concept that describes a probabilistic framework for recursive state estimation. Its main idea consists in estimating the distribution of an uncertainty, called the *belief*, over the possible state space (e.g., the range of possible user kinetics) given the sensor measurements obtained up to the time of the estimation. If we define t_k as the instant in time when the kth measurement is obtained, x_k as the random variable that represents the state space at time t_k and $Z_{1:k}$ as the set of all sensor measurements until time t_k, the *belief* is defined as:

$$\mathrm{Bel}(x_k) = \mathrm{f}(x_k|Z_{1:k}), \tag{5.66}$$

where $\mathrm{f}(x_k|Z_{1:k})$ is the probability density function of x_k given $Z_{1:k}$. Note that in positioning applications, x_k often represents the user kinetics. However, in this section, x_k is considered as an arbitrary state space.

As Equation (5.66) shows, the belief has a dependency on all measurements obtained from time t_0 until the present time t_k, meaning that the complexity increases with time. To solve the problem presented by this dependency, it is assumed that the Bayesian filter has Markovian properties of the first order, i.e., all the information necessary to compute an estimation of position at time t_k is given by the information available at time t_{k-1}.

The Bayesian filter has two main steps:

- **Prediction.** At each time, the current state x_k is updated based on the previous state x_{k-1}:

$$\mathrm{Bel}^{-}(x_k) = \int \mathrm{f}(x_k|x_{k-1})\mathrm{Bel}(x_{k-1})\,\mathrm{d}x_{k-1}. \tag{5.67}$$

  In Equation (5.67), the term $\mathrm{f}(x_k|x_{k-1})$ describes the *system dynamics*, i.e., how the system evolves from the state x_{k-1} to the current state x_k. The belief at time t_0 is usually defined as a reasonable density function within the context of the system in use.

- **Correction.** The predicted belief is corrected every time a new sensor measurement is obtained:

$$\mathrm{Bel}(x_k) = \varrho_k \mathrm{f}(z_k|x_k)\mathrm{Bel}^{-}(x_k). \tag{5.68}$$

  In Equation (5.68), the term ϱ_k represents the normalization constant that is used in order to ensure that $\int \mathrm{Bel}(x_k)\,\mathrm{d}x_k = 1$. The term $\mathrm{f}(z_k|x_k)$ represents the *observation model*, i.e. the probability of having a sensor measurement z_k given that the system is in state x_k.

For localization purposes, the term $\mathrm{f}(x_k|x_{k-1})$ in Equation (5.67) is determined by the motion model, while the term $\mathrm{f}(z_k|x_k)$ in Equation (5.68) is dependent on the technology of the sensor and the location method. For instance, the latter term could be the path loss expression of Section 3.2.1 if we are considering an RSS-based positioning system.

5.3.1 The Kalman filter

The Kalman filter (Grewal and Andrews 2001; Kailath et al. 2000; Kalman 1960) is one of the most widely used implementations of the Bayesian filter. The basic assumptions of this filter are that the models $f(x_k|x_{k-1})$ in Equation (5.67) and $f(z_k|x_k)$ in Equation (5.68) are linear, and that the noise statistics have a Gaussian distribution, i.e., the belief is Gaussian. The filter defines five equations (set of expressions given by Equations (5.73)–(5.77)), which, in a recursive manner, can estimate the state of an unknown noisy process. It is a very powerful filter because the amount of processing is considerably reduced when compared with other Bayesian filters (such as the particle filter described in Section 5.3.2) and because it estimates past, present and even future states using approximations to the true system model. Since this filter was proposed, it has been successfully applied in several areas of research. An example of such an application is in the tracking of moving devices.

The Kalman filter equations are able to predict the state vector X of a discrete-time process that is governed by the following linear system in state space form:

$$X_k = AX_{k-1} + BU_{k-1} + \mathbf{w}_{k-1}, \quad \zeta \sim \text{Norm}(0,\, Q), \tag{5.69}$$

$$Z_k = HX_k + \mathbf{v}_k, \quad \eta \sim \text{Norm}(0,\, R). \tag{5.70}$$

In Equation (5.69), the matrix A defines the relation between the previous state and the current one, the matrix B relates the previous external input vector U_k to the current state, and ζ is a random variable representing the process noise. The control input U can be either known or unknown. In Equation (5.70), the matrix H represents the relation between the present state X_k and the measurement Z_k expected for that state X_k. The random variable η represents the measurement noise. In Equations (5.69) and (5.70), the random variables ζ and η are assumed to be independent of each other, to be white and to have a normal distribution with zero mean and covariance matrices Q and R, respectively. Having defined the matrices in Equations (5.69) and (5.70) and the related parameters, we still need to define the starting values for the state $X_{0|0}$ and the error covariance estimator $P_{0|0}$. These parameters are system-dependent. For the state $X_{0|0}$, this value is the expected value of the state space at time t_0. The estimator $P_{0|0}$ is commonly initialized with the covariance matrix Q. Based on the system of Equations (5.69) and (5.70), the Kalman filter can be defined as:

Initialization:

$$\hat{X}_{0|0} = \text{E}[X_0], \tag{5.71}$$

$$P_{0|0} = \text{E}[(\hat{X}_{0|0} - X_0)(\hat{X}_{0|0} - X_0)^{\text{T}}] \simeq Q. \tag{5.72}$$

Prediction:

$$\hat{X}_{k|k-1} = A_k \hat{X}_{k-1|k-1} + B_k U_{k-1|k-1}, \tag{5.73}$$

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^{\text{T}} + Q_k. \tag{5.74}$$

Correction:

$$K_k = P_{k|k-1} H_k^{\mathrm{T}} (H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k)^{-1}, \tag{5.75}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k (Z_k - H_k \hat{X}_{k|k-1}), \tag{5.76}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}. \tag{5.77}$$

From Equations (5.75) and (5.76), the "innovation process" operating in this context can be identified as the difference between the obtained measurements Z_k and the expected measurements $H_k \hat{X}_{k|k-1}$. This process, by definition Gaussian-distributed, defines how much each individual measurement Z_k contributes to the overall estimation of the state space $\hat{X}_k$. The innovation process and its associated covariance are defined as:

$$\tilde{Z}_k = Z_k - H_k \hat{X}_{k|k-1}, \tag{5.78}$$

$$S_k = H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k. \tag{5.79}$$

In Section 6.4.3.1, this innovation process is used as a solution for estimating abrupt changes in the dynamics of the MS, often referred to as "maneuvers".

As an alternative form of the KF for scenarios where the positioning system is not run in real time, we shall now present the fixed-lag Kalman smoother (Cohn et al. 1994; Todling et al. 1997). This smoother runs a Kalman filter similar to the one presented above in this section and then a set of additional equations called (Todling et al. 1997) *retrospective-analysis* equations. A smoothed estimation $\hat{X}_{k-\ell|k}$ at a past time $t_{k-\ell}$, $\ell > 0$, can be obtained by running the expressions of the ordinary Kalman filter until time $t_{k-\ell}$ and then, beyond that time until the current time t_k, running the *retrospective-analysis* equations. The smoothed estimator $\hat{X}_{k-\ell|k}$ of the hidden process beyond time $t_{k-\ell}$, given observations until time t_k, is calculated from:

Smoothing:

$$\hat{X}_{k-\ell|k} = \hat{X}_{k-\ell|k-1} + K_{k-\ell|k} \tilde{Z}_k, \tag{5.80}$$

$$K_{k-\ell|k} = \left(P^{\mathrm{fa}}_{k,k-\ell|k-1}\right)^{\mathrm{T}} H_k^{\mathrm{T}} S_k^{-1}, \tag{5.81}$$

$$P^{\mathrm{fa}}_{k,k-\ell|k-1} = A_k P^{\mathrm{aa}}_{k-1,k-\ell|k-1}, \tag{5.82}$$

$$P^{\mathrm{aa}}_{k,k-\ell|k} = P^{\mathrm{fa}}_{k,k-\ell|k-1} - K_k H_k P^{\mathrm{fa}}_{k,k-\ell|k-1}, \tag{5.83}$$

where the quantities $\tilde{Z}_k$ and S_k are given by Equations (5.78) and (5.79), respectively. The matrix $P^{\mathrm{fa}}_{k,k-\ell|k-1}$ is called the forecast retrospective-analysis error cross-covariance matrix and $P^{\mathrm{aa}}_{k,k-\ell|k}$ is the filter analysis retrospective-analysis error cross-covariance matrix. Basically, the matrices $P^{\mathrm{fa}}_{k,k-\ell|k-1}$ and $P^{\mathrm{aa}}_{k,k-\ell|k}$ propagate along

time the uncertainty in the estimation at time $t_{k-\ell}$, given the measurements until time t_k. The covariance error $P_{k-\ell|k}$ at time $t_{k-\ell}$ is given by:

$$P_{k-\ell|k} = P_{k-\ell|k-1} - K_{k-\ell|k} H_k P^{\text{fa}}_{k,k-\ell|k-1}. \tag{5.84}$$

The set of Equations (5.80)–(5.84), together with the Kalman filter Equations (5.73)–(5.77), define the Kalman smoother.

5.3.1.1 Extended Kalman filter

As we have seen in the previous section, the Kalman filter is an algorithm that requires the models to be linear. The main problem with this approach is that in typical cases where such an algorithm could be useful, the models of the system are nonlinear, i.e.,

$$X_k = \mathcal{F}(X_{k-1}, U_{k-1}, \zeta_{k-1}), \quad \zeta \sim \text{Norm}(0, Q), \tag{5.85}$$

$$Z_k = \mathcal{H}(X_k, \eta_k), \quad \eta \sim \text{Norm}(0, R). \tag{5.86}$$

Researchers have proposed several extensions of the Kalman filter for nonlinear cases. The most used solutions are the extended Kalman filter (EKF), described in detail for example by Grewal and Andrews (2001), and the unscented Kalman filter (UKF), proposed by Julier and Uhlmann (1997). These solutions deal with the nonlinearity by approximating it with a Taylor expansion. While the EKF approximates the models with their first derivative, the UKF approximates them with their second derivative.

The EKF is a commonly used solution when the nonlinearity is not "very" strong, meaning that the error in this approximation is several orders of magnitude lower than the typical errors introduced by the noise in the system. The solution uses a Taylor expansion truncated the linear terms, resulting in expressions for the models equivalent to Equations (5.69) and (5.70). The difference is that the terms A and H are now Jacobian matrices instead of precise relationships. These Jacobian matrices can be defined as:

$$A = \frac{\partial \mathcal{F}}{\partial X}(X_{k-1}, U_{k-1}, 0), \tag{5.87}$$

$$H = \frac{\partial \mathcal{H}}{\partial X}(X_k, 0). \tag{5.88}$$

Based on the nonlinear Equations (5.85) and (5.86) and the parameters in Equations (5.87) and (5.88), the EKF equations can be defined as:

Initialization:

$$\hat{X}_{0|0} = \text{E}[X_0], \tag{5.89}$$

$$P_{0|0} = \text{E}[(\hat{X}_{0|0} - X_0)(\hat{X}_{0|0} - X_0)^{\text{T}}] \simeq Q. \tag{5.90}$$

Prediction:

$$\hat{X}_{k|k-1} = \mathcal{F}(\hat{X}_{k-1|k-1}, U_{k-1|k-1}, 0), \tag{5.91}$$

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^{\mathrm{T}} + Q_k. \tag{5.92}$$

Correction:

$$K_k = P_{k|k-1} H_k^{\mathrm{T}} (H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k)^{-1}, \tag{5.93}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k (Z_k - \mathcal{H}(\hat{X}_{k|k-1}, 0)), \tag{5.94}$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}. \tag{5.95}$$

One of the drawbacks of the EKF approach for stochastic filtering is, as previously stated, the linearization of the models. This linearization, depending on the application, can represent a high impact on the system concerning the errors in the estimated mean and covariance. The problem becomes more and more evident as higher-order terms become more important in the nonlinear model. To provide a more accurate approximation of the models, the UKF was proposed.

5.3.1.2 Unscented Kalman filter

Similarly to the EKF, the UKF assumes the noise in the models to follow a Gaussian random distribution. The difference is that the distributions are represented by a minimal set of sample points chosen in such a way that the true mean and covariance of the errors are captured. As was shown by Julier and Uhlmann (1997), these samples, when propagated through the nonlinearity, accurately capture the posterior mean and covariance up to the third-order Taylor expansion of the nonlinearity. This procedure is known as the unscented transformation. The algorithm augments the hidden space X_k with noise components such that the hidden space becomes $X^a = [X^{\mathrm{T}} \quad \zeta^{\mathrm{T}} \quad \eta^{\mathrm{T}}]^{\mathrm{T}}$. Then the algorithm can be described as:

Initialization:

$$\hat{X}_{0|0} = \mathrm{E}[X_0], \tag{5.96}$$

$$P_{0|0} = \mathrm{E}[(\hat{X}_{0|0} - X_0)(\hat{X}_{0|0} - X_0)^{\mathrm{T}}] \simeq Q, \tag{5.97}$$

$$\hat{X}^a_{0|0} = \mathrm{E}[X^a_0] = [\hat{X}^{\mathrm{T}}_{0|0} \quad \mathbf{0} \quad \mathbf{0}]^{\mathrm{T}}, \tag{5.98}$$

$$P^a_{0|0} = \mathrm{E}[(\hat{X}^a_{0|0} - X^a_0)(\hat{X}^a_{0|0} - X^a_0)^{\mathrm{T}}] = \begin{bmatrix} P_{0|0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & R \end{bmatrix}, \tag{5.99}$$

where $\mathbf{0}$ is a zero matrix of appropriate size, and Q and R are the noise covariance matrices of the hidden process (in Equation (5.85)) and observable process (in Equation (5.86)), respectively. The prediction phase of the algorithm is:

Prediction:

$$\mathsf{X}^a_{k-1|k-1} = \left[\hat{X}^a_{k-1|k-1} \quad \hat{X}^a_{k-1|k-1} \pm \sqrt{(L+\lambda)P^a_{k-1|k-1}}\right], \tag{5.100}$$

$$\mathsf{X}^x_{k|k-1} = \mathcal{F}(\mathsf{X}^x_{k-1|k-1}, 0, \mathsf{X}^w_{k-1|k-1}), \tag{5.101}$$

$$\hat{X}_{k|k-1} = \sum_{i=0}^{2L} W_i^{(m)} \mathsf{X}^x_{k|k-1;i}, \tag{5.102}$$

$$P_{k|k-1} = \sum_{i=0}^{2L} W_i^{(c)} [\mathsf{X}^x_{k|k-1;i} - \hat{X}_{k|k-1}][\mathsf{X}^x_{k|k-1;i} - \hat{X}_{k|k-1}]^{\mathrm{T}}, \tag{5.103}$$

$$\mathsf{Z}_{k|k-1} = \mathcal{H}(\mathsf{X}^x_{k|k-1}, \mathsf{X}^v_{k|k-1}), \tag{5.104}$$

$$\hat{Z}_{k|k-1} = \sum_{i=0}^{2L} W_i^{(m)} \mathsf{Z}_{k|k-1;i}. \tag{5.105}$$

In Equation (5.100), $\mathsf{X}^a = [(\mathsf{X}^x)^{\mathrm{T}} \quad (\mathsf{X}^w)^{\mathrm{T}} \quad (\mathsf{X}^v)^{\mathrm{T}}]$ is the sigma points, i.e., the points where the noise distributions are sampled. The notation $\sqrt{(L+\lambda)P^a_{k-1|k-1}}$ corresponds to the Cholesky factorization, and its summation with $\hat{X}^a_{k-1|k-1}$ corresponds to a columnwise sum. The parameters L and λ are the dimension of X and a scaling parameter, respectively; the latter is set to $\lambda = (10^3)^2 L - L$ as advised in Julier and Uhlmann (1997). In Equations (5.101) and (5.104), $\mathcal{F}$ and $\mathcal{H}$ are, respectively, the evolution model and the observation model as defined by Equations (5.85) and (5.86). In Equations (5.102) and (5.105), the weights $W_i^{(m)}$ are calculated as in Julier and Uhlmann (1997) and the symbol $\mathsf{X}^x_{k|k-1;i}$ corresponds to column i of the matrix $\mathsf{X}^x_{k|k-1}$. The correction phase of the algorithm is:

Correction:

$$P_{z_k z_k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathsf{Z}_{k|k-1;i} - \hat{Z}_{k|k-1}][\mathsf{Z}_{k|k-1;i} - \hat{Z}_{k|k-1}]^{\mathrm{T}}, \tag{5.106}$$

$$P_{x_k z_k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathsf{X}_{k|k-1;i} - \hat{X}_{k|k-1}][\mathsf{Z}_{k|k-1;i} - \hat{Z}_{k|k-1}]^{\mathrm{T}}, \tag{5.107}$$

$$K_k = P_{x_k z_k} P^{-1}_{z_k z_k}, \tag{5.108}$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Z_k - \hat{Z}_{k|k-1}), \tag{5.109}$$

$$P_{k|k} = P_{k|k-1} + K_k P_{z_k z_k} K_k^{\mathrm{T}}, \tag{5.110}$$

where the $W_i^{(c)}$, similarly to the $W_i^{(m)}$, are calculated as in Julier and Uhlmann (1997). The matrix $P_{z_k z_k}$ corresponds to the innovation covariance and $P_{x_k z_k}$ corresponds to the cross-correlations between the hidden and the observable process.

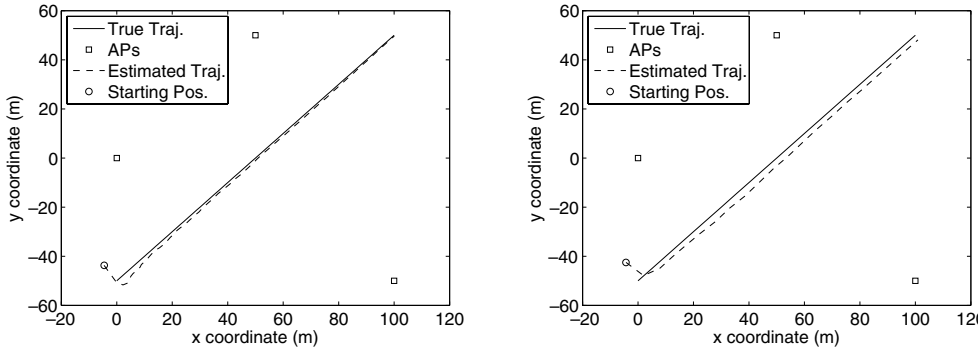


Figure 5.2 Example of an application where a wireless target is tracked using three access points capable of measuring the distance to the target. The left plot shows the tracking result when fresh measurements are available every time the target moves of 1 m. The right plot shows the equivalent result when fresh measurements are available every time the target moves of 5 m.

5.3.1.3 Convergence issues

Iterative methods such as the Kalman filter have a particular aspect concerning the convergence of the estimators that cannot be disregarded. This aspect is connected to the number of steps of the various iterations that are performed. In the case of the Kalman filter, since each step is performed whenever fresh measurements are obtained, the convergence issue is tightly connected to the frequency of fresh measurements.

Let us assume a simple scenario where three access points (APs) measure the distance to a target that is moving through their coverage range. We consider two cases, where fresh measurements are available every time the target moves either 1 m or 5 m. We have run a filter for these two cases with the appropriate parameters[5]. Figure 5.2 shows the sequence of position estimators when the extended Kalman filter is used. As can be seen, more frequent measurements (Figure 5.2 (left) in comparison with Figure 5.2 (right)) means faster convergence to the true trajectory. The same conclusion can be drawn by calculating the Euclidean distance between the true position and the estimated position for every time step. Figure 5.3 shows that this Euclidean distance tends to lower the values as more measurements are available. At this point, however, it is important to mention that this tendency may not be necessarily true if there are non-Gaussian errors in the measurements or in the presence of colored noise.

Though the Euclidean distance between the true and the estimated position indicates the estimation error, the true trajectory is seldom known. For this reason, it is common to analyze the trace of the covariance matrix P. As Figure 5.4 shows, the

[5]Since we analyze the influence of measurement frequency and not the influence of filter configuration in this section, the details of the filter are not given here.

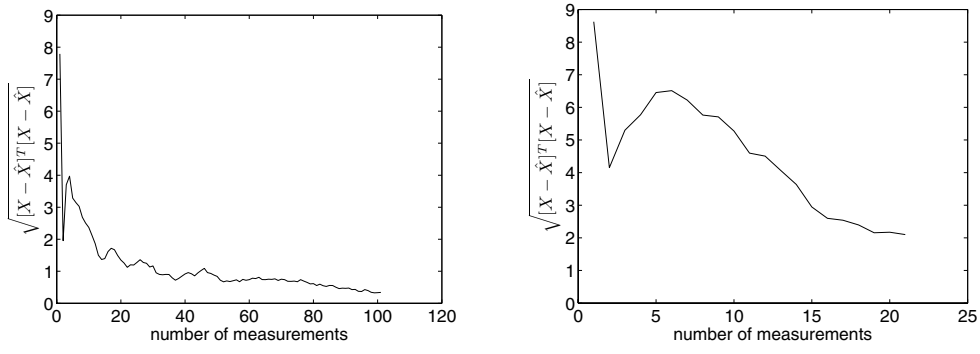


Figure 5.3 Evolution of the Euclidean distance between the true position and the estimated position at every time step. The left plot was obtained for a setup where fresh measurements are available every time the target moves of 1 m. The right plot was obtained for a setup where measurements are available every time the target moves of 5 m.

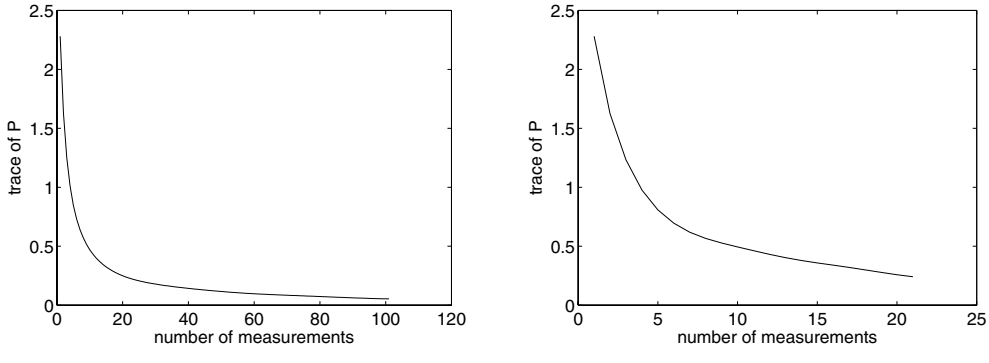


Figure 5.4 Evolution of the trace of the covariance matrix P at every time step. The left plot was obtained in a setup where fresh measurements are available every time the target moves of 1 m. The right plot was obtained for a setup where measurements are available every time the target moves of 5 m.

trace of P tends to become lower and lower as the number of available measurements increases. This value gives a level of confidence for the estimator of position. Thus, a common rule to define the convergence of estimators is to define a threshold below which the filter is said to converge.

5.3.2 The particle filter

The "particle filter" (PF) is the name given to a Monte Carlo implementation of the Bayesian framework. This filter is adequate for nonlinear and non-Gaussian

estimation; however, it suffers from the numerical problems that characterize Monte Carlo algorithms. Several different approaches are used, depending on characteristics of the noisy set of data. As stated by Arulampalam et al. (2002), the algorithms used have a basis in sequential importance sampling. From the available solutions, we present here the "bootstrap filter" introduced by Gordon et al. (1993).

Suppose that at time t_{k-1}, M random samples $X^{(m)}_{k-1|k-1}$ $(1 \leq m \leq \mathsf{M})$, commonly referred to as particles, are available and that they are drawn from a probability distribution $\mathrm{f}(X_{k-1}|Z_{1:k})$, i.e., the belief $\mathrm{Bel}(X_{k-1})$, as expressed in Equation (5.66). The bootstrap filter is an algorithm that propagates these samples through the system models (i.e., the models of Equations (5.85) and (5.86)) in order to obtain random samples $X^{(m)}_{k|k}$ that approximate the belief $\mathrm{Bel}(X_k)$ at time t_k. Similarly to any Bayesian filter, the bootstrap filter runs in a cyclic fashion:

Initialization:

$$X^{(m)}_{0|0} \sim \mathrm{f}(X_0), \tag{5.111}$$

Prediction:

$$X^{(m)}_{k-1|k-1} \sim \mathrm{Bel}(X_{k-1}), \tag{5.112}$$

$$X^{(m)}_{k|k-1} = \mathcal{F}(X^{(m)}_{k-1|k-1}, U_{k-1|k-1}, \zeta^{(m)}_{k-1}) \quad \zeta^{(m)}_{k-1} \sim \mathrm{f}(\zeta_{k-1}), \tag{5.113}$$

where the $\zeta^{(m)}_{k-1}$ are samples drawn from the noise component ζ in the system dynamics;

Correction:

$$W^{(m)}_k = \frac{\mathrm{f}(Z_k|X^{(m)}_{k|k-1})}{\sum_{m=1}^{\mathsf{M}} \mathrm{f}(Z_k|X^{(m)}_{k|k-1})}, \quad \mathrm{f}(Z_k|X^{(m)}_{k|k-1}) = \int \delta(Z_k - \mathcal{H}(X^{(m)}_{k|k-1}, \eta_k))\mathrm{f}(\eta_k)\,\mathrm{d}\eta_k, \tag{5.114}$$

$$\hat{X}_{k|k} = \sum_{m=1}^{\mathsf{M}} W^{(m)}_k X^{(m)}_{k|k-1}, \tag{5.115}$$

$$\mathrm{Bel}(X_k) = \sum_{m=1}^{\mathsf{M}} W^{(m)}_k \delta(X_k - X^{(m)}_{k|k-1}), \tag{5.116}$$

where $\hat{X}_{k|k}$ is the estimator of $X_{k|k}$ at time t_k given measurements until time t_k.

5.3.3 Grid-based methods

The grid-based methods are optimal implementations of the Bayesian framework when the state space is discrete and finite. Suppose that the state space at time $k-1$ consists of discrete states $X^{(i)}_{k-1|k-1}$, with $i = 1, \ldots, l_s$. For each state $X^{(i)}_{k-1|k-1}$, let

the belief of that state, i.e., the conditional probability of that state given observation until time $k-1$, be denoted by $w^{(i)}_{k-1|k-1}$; that is, $\text{Bel}(X_{k-1} = X^{(i)}_{k-1|k-1}) = \Pr(X_{k-1} = X^{(i)}_{k-1|k-1}|Z_{1:k-1}) = w^{(i)}_{k-1|k-1}$. Then the a posteriori PDF at $k-1$ can be written as:

$$\text{Bel}(X_{k-1}) = \sum_{i=1}^{l_s} w^{(i)}_{k-1|k-1}\delta(X_{k-1} - X^{(i)}_{k-1|k-1}), \tag{5.117}$$

where $\delta()$ is the Dirac function. Using Equation (5.117) in the update and in the correction function, it is possible to define the grid-based filter by the following equations:

Initialization:

$$\hat{X}_{0|0} = \arg\max_{X^{(i)}_{0|0}}(w^{(i)}_{0|0}), \tag{5.118}$$

Prediction:

$$w^{(i)}_{k|k-1} = \sum_{j=1}^{l_s} w^{(j)}_{k-1|k-1} P(X^{(i)}_{k|k}|X^{(j)}_{k-1|k-1}), \tag{5.119}$$

$$\text{Bel}^{-}(X_k) = \sum_{i=1}^{l_s} w^{(i)}_{k|k-1}\delta(X_k - X^{(i)}_{k|k}), \tag{5.120}$$

Correction:

$$w^{(i)}_{k|k} = \frac{w^{(i)}_{k|k-1} P(Z_k|X^{(i)}_{k|k})}{\sum_{j=1}^{l_s} w^{(j)}_{k|k-1} P(Z_k|X^{(j)}_{k|k})}, \tag{5.121}$$

$$\text{Bel}(X_k) = \sum_{i=1}^{l_s} w^{(i)}_{k|k}\delta(X_k - X^{(i)}_{k|k}), \tag{5.122}$$

$$\hat{X}_{k|k} = \arg\max_{X^{(i)}_{k|k}}(w^{(i)}_{k|k}). \tag{5.123}$$

The above equations require the distributions $p(X^{(i)}_{k|k}|X^{(j)}_{k-1|k-1})$ and $p(Z_k|X^{(j)}_{k|k})$ to be known, although they are not constrained to be any particular form of distribution. Note that in Equations (5.118) and (5.123), the estimated state is defined as the maximum likelihood of the discrete belief.

5.4 Estimating Model Parameters and Biases in Observations

In wireless positioning and in general in any estimation process, it can happen that some of the model parameters are unknown or even that the models are unknown. In this section, we present a few typical methods for estimating such unknown parameters.

5.4.1 Precalibration

A very simple but time-consuming method for determining model parameters and biases in the measurements is to run a precalibration phase. In this phase, typical MS movements are performed within a test area in order to obtain actual measurements of the positioning system. For instance, in a WLAN positioning solution, one would need to obtain RSS measurements. While the true position of the MS gives us the so-called ground truth, the measurements give us data to estimate the target position. Combined analysis of the two types of data gives us information on the missing parameters.

A typical, generic method of precalibration might run as follows:

- Typical target trajectories are defined, and both true position data and corresponding timestamps are recorded. This data defines the ground truth.
- During the movement of the target, the system obtains measurements from the radio links that are available and are being calibrated. The timestamps of the measurements are also stored.
- By using appropriate models, the data measured from the radio links is translated into position information. To accomplish this, the unknown parameters of the models must be chosen on a rule-of-thumb basis.
- By comparing the estimated position information and ground truth, the parameters are determined either by trial-and-error tests or by using more advanced mathematical approaches.

Note, however, that the number of parameters to be estimated must guarantee the observability of the system. Only in this case can a single solution be found for each parameter.

Let us assume that the ground truth position at time k is given by X_k^{true}. We assume also that at that same time k, the measurement Z_k was obtained and that the measurements have a linear relation with the parameters ξ to be estimated:

$$Z = H\xi + \eta, \tag{5.124}$$

where each H_k is a function of the ground truth X_k^{true}. Using the LS estimator, the parameters can be determined from:

$$\hat{\xi} = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}Z, \tag{5.125}$$

where Z and H stack all the measurements Z_k and terms H_k along time into column vectors. Then, based on the estimator $\hat{\xi}$, it is possible to determine the covariance matrix of the parameters from:

$$P = (H^{\mathrm{T}}R^{-1}H)^{-1}, \tag{5.126}$$

with $R = \mathrm{E}[\{Z - H\hat{\xi} - \mathrm{E}[Z - H\hat{\xi}]\}\{Z - H\hat{\xi} - \mathrm{E}[Z - H\hat{\xi}]\}^{\mathrm{T}}] = \mathrm{E}[\eta\eta^{\mathrm{T}}]$ with η zero mean distributed.

5.4.2 Joint parameter and state estimation

In joint estimation, both the state space and the parameters are estimated together. In order to actually implement such a concept, it is necessary to augment Equations (5.69) and (5.70) with the parameters:

$$\left[\begin{array}{c} X_k \\ \hline \xi_k \end{array}\right] = \left[\begin{array}{c|c} A & A_\xi \\ \hline 0 & 1 \end{array}\right] \left[\begin{array}{c} X_{k-1} \\ \hline \xi_{k-1} \end{array}\right] + \left[\begin{array}{c} B \\ \hline 0 \end{array}\right] U_{k-1} + \zeta_{k-1}, \qquad \zeta \sim \text{Norm}\left(0, \left[\begin{array}{c|c} Q & 0 \\ \hline 0 & Q_\xi \end{array}\right]\right), \tag{5.127}$$

$$Z_k = \left[H \mid 0\right] \left[\begin{array}{c} X_k \\ \hline \xi_k \end{array}\right] + \eta_k \quad \text{and} \quad \eta \sim \text{Norm}\left(0, \left[\begin{array}{c|c} R & 0 \\ \hline 0 & R_\xi \end{array}\right]\right), \tag{5.128}$$

where the parameters ξ are considered independent of time. Using the new definition of the state space and observation vector, the desired estimation algorithm can be used.

5.5 Alternative Approaches

Though LS algorithms and Bayesian filtering techniques are some of the most used techniques to circumvent errors in the measurements obtained from radio channels, several other options exist in the literature.

5.5.1 Fingerprinting

Let us assume that for localizing an MS, measurements from the communication channel are obtained by several BSs identified by an index i. Let us further assume a general case where a precalibration phase is performed. The precalibration is done in such a way that for scattered positions X over the entire scenario, the distribution of the observed measurements Z at BS_i is determined as $\text{f}(Z^{(i)}|X)$. This information is considered as the calibration content of a database. Let us now suppose that new data $Z^{(i)}_{\text{new}}$ is available in order to perform positioning:

$$\text{f}(X|Z^{(i)}_{\text{new}}) = \frac{\text{f}(Z^{(i)}_{\text{new}}|X)\text{f}(X)}{\text{f}(Z^{(i)}_{\text{new}})}. \tag{5.129}$$

Given that $\text{f}(Z^{(i)}_{\text{new}})$ is not straightforward to determine,

$$\text{f}(X|Z^{(i)}_{\text{new}}) = \frac{\text{f}(Z^{(i)}_{\text{new}}|X)\text{f}(X)}{\int_X \text{f}(Z^{(i)}_{\text{new}}|X)\text{f}(X)\,\text{d}X}. \tag{5.130}$$

Assuming that the measurements from the various APs are i.i.d. when conditioned to X,

$$\mathrm{f}(X|Z_{\mathrm{new}}) = \frac{\mathrm{f}(Z_{\mathrm{new}}^{(i)}|X)\mathrm{f}(X)}{\int_X \mathrm{f}(Z_{\mathrm{new}}^{(i)}|X)\mathrm{f}(X)\,\mathrm{d}X}. \tag{5.131}$$

Note that in Equation (5.131) the density function $\mathrm{f}(Z_{\mathrm{new}}^{(i)}|X)$ is obtained from the calibration information contained in the database, i.e., the density value is obtained as $\mathrm{f}(Z^{(i)} = Z_{\mathrm{new}}^{(i)}|X)$. Given Equation (5.131), it is possible to estimate the position of the MS using the maximum-likelihood estimator:

$$\hat{X} = \arg\max_{X|P}(\mathrm{f}(X|Z_{\mathrm{new}})). \tag{5.132}$$

As an example, a simulation where three APs measure received-power information has been performed. The MS was assumed to be static, continuously communicating with all the APs, and the measurements of power were assumed to be corrupted by Gaussian noise. Since the purpose of this simulation was not to provide details of the actual implementation and positioning accuracy but instead to illustrate the methodologies analyzed in this section, the values of the parameters will not be mentioned here. The calibration phase was performed over the entire 10 m $\times$ 10 m area under consideration using a grid with a granularity of 5 cm. Figure 5.5 shows the expected value of $\mathrm{f}(Z^{(i)}|X)$ and the corresponding 1σ confidence intervals for each AP. After the calibration, the MS was placed in a random position and new measurements of received power were obtained. Using the new measurements and the calibration data, Equations (5.130) and (5.131) were applied, and the results are plotted in Figures 5.6 and 5.7, respectively. As can be seen in Figure 5.6, the positions with equal density of probability are arranged in a circular shape. This fact is a consequence of the dependency that received-power measurements have on the distance between the AP and the MS. In Figure 5.7 it is possible to see that the density of probability given by Equation (5.131), which takes into account all APs, shows a cone-like shape. This shape is a consequence of the product of the distributions in Figure 5.6. Note, however, that as stated in Section 4.6.2 the distribution shown in Figure 5.7 is highly dependent on the placement of the APs. The estimation of position resulting from Equation (5.132) is shown in Figure 5.8. In contrast to the LS and Bayesian approaches, fingerprinting techniques are typically more accurate in scenarios with several physical obstructions in the communication links (e.g., indoor scenarios). The disadvantage of fingerprinting techniques is that they depend strongly on the preacquired data, resulting in weakness when the propagation conditions change.

5.5.2 Time series data

Exponential smoothing techniques are techniques often used in the context of business and economics (Box and Jenkins 1976; NIST/SEMATECH 2007). Their application in the context of positioning is possible and has been used in previous

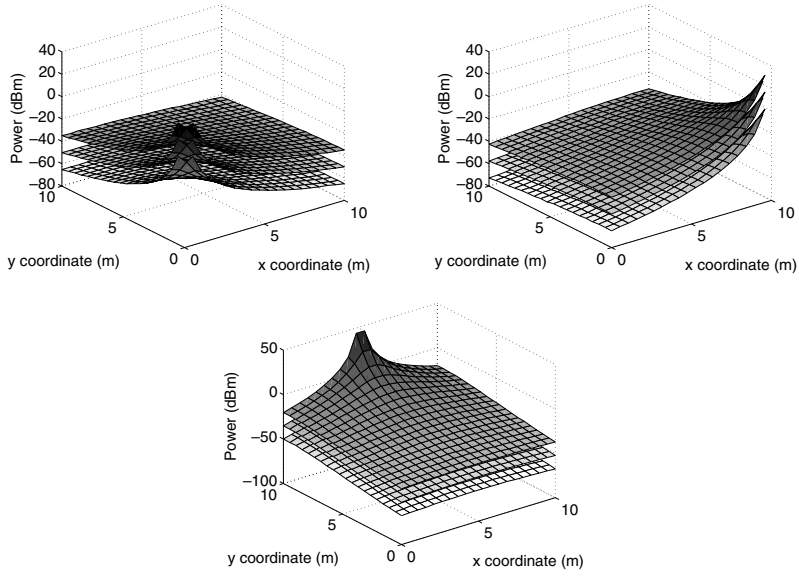


Figure 5.5 Each of these three plots shows the average and the 1σ confidence interval as a function of position. The top curve shows the top limit of the 1σ confidence interval, the middle curve shows the mean value and the bottom curve shows the bottom limit of the 1σ confidence interval. Each plot corresponds to a different AP.

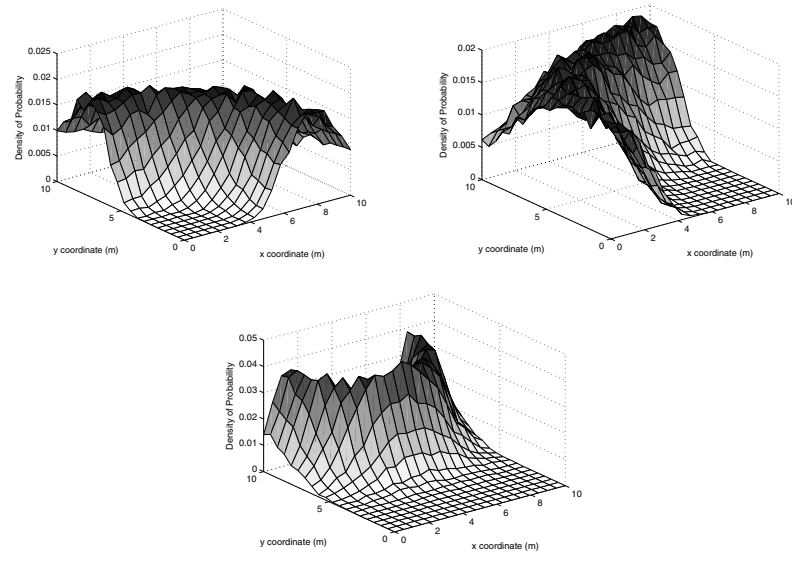


Figure 5.6 These plots show, for each AP, the distribution of probability given by Equation (5.130), i.e., the probability of placement of the MS given the measurements of power and the entire calibration data.

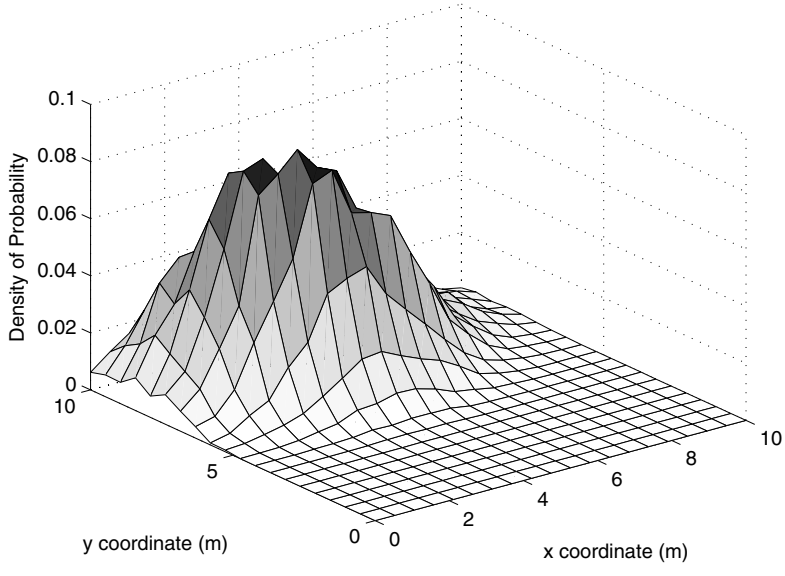


Figure 5.7 Probability density function of placement of the MS with respect to the measurements obtained.

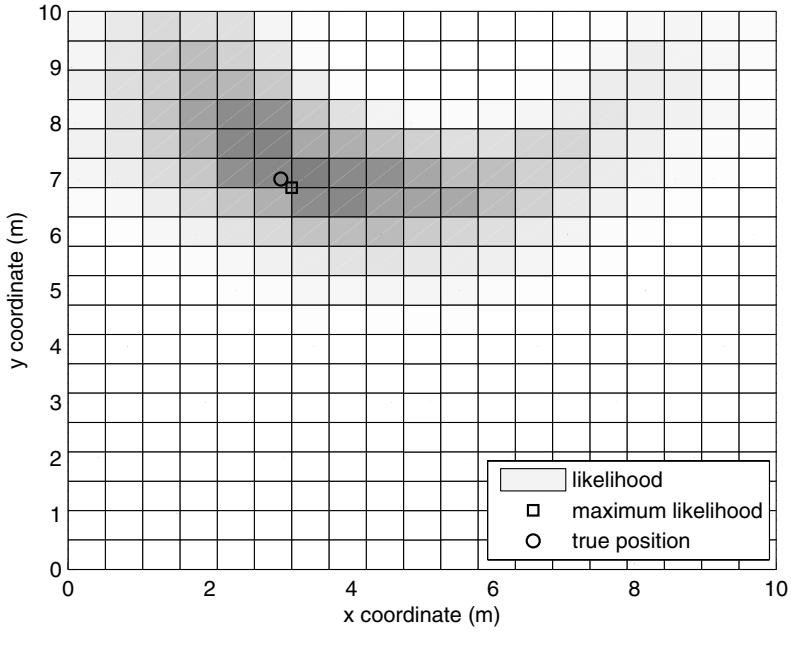


Figure 5.8 Heat map representing the likelihood of the MS position. The darker the map, the more likely is the placement of the MS.

work (Aloi and Korniyenko 2007; LaViola 2003). Time series techniques suit applications where data points, corrupted by noise, have an internal structure such as trends or seasonal variations.

This type of techniques is simpler to implement than the techniques described above; however, they are mostly suitable for systems where the MS position has been estimated beforehand. In this case, time series are appropriate tools for smoothing the variations in a sequence of position estimations.

5.5.2.1 Single exponential smoothing

The single exponential smoothing technique is the simplest technique of the family of exponential smoothing (NIST/SEMATECH 2007). This technique uses an exponential weighted averaging law in order to put more emphasis on more recent observations. This technique assumes no trends, and thus it is most appropriate for localizing static nodes or nodes moving at velocities low enough that the position can be considered constant over all the averaging time window.

Thus, supposing that at time t_k an observation of the position X_k of the MS is available, the smoothed estimator of the location of the MS is given by:

$$S_k = \gamma_s X_k + (1 - \gamma_s) S_{k-1}, \quad 0 \leq \gamma_s \leq 1, \tag{5.133}$$

where the variable γ_s is the smoothing forgetting factor. In the absence of other information, the initial condition is commonly set as $S_1 = X_1$. The observations of position X_k can be obtained by using the models presented in Chapter 4.

5.5.2.2 The double exponential smoother

When the data points show a trend, a second exponential smoothing equation can be used for estimating that trend. The idea is to consider a second equation that estimates the constant increment that each data point is subject to with time:

$$\begin{cases} S_k = \gamma_s X_k + (1 - \gamma_s)(S_{k-1} + B_{k-1}), & 0 \leq \gamma_s \leq 1, \\ B_k = \gamma_b (S_k - S_{k-1}) + (1 - \gamma_b) B_{k-1}, & 0 \leq \gamma_b \leq 1, \end{cases} \tag{5.134}$$

where the parameter γ_b is the forgetting factor of the incremental trend B_k. In practice, S_k is the smoothed estimator of the position data points and B_k is the smoothed estimator of the increment that the data points are subject to at each time step. It is important to note that the first line of Equation (5.134) includes the trend B_{k-1} at a previous time, meaning that it takes into account the time lag between two consecutive data points. The initial conditions of Equation (5.134) are commonly set as:

$$\begin{cases} S_1 = X_1, \\ B_1 = X_2 - X_1. \end{cases} \tag{5.135}$$

Forecasting the data points into the future is possible using the following equation:

$$S_{k+m}^{\mathrm{f}} = S_k + mB_t, \tag{5.136}$$

where m is the number of time steps that the forecast is looking at.

5.6 Conclusions

This chapter has focused on data fusion techniques for positioning applications. The traditional methods for data fusion were described in the context of wireless positioning. The LS methods were described as simple mathematical formulations to be used primarily in static scenarios. Owing to the necessity for tracking moving users, the Bayesian framework was analyzed. Within this framework, the main focus was given to Kalman filters, and less focus on particle filters and grid-based methods. Kalman filters are studied as convenient algorithms for tracking MSs in the presence of white Gaussian noise. In contrast, particle filters are suitable for any kind of noise; however, the computational complexity may be considerably higher compared with KF methods. Grid-based methods are used when the state space, i.e., the possible positions for finding the MS, is discrete and finite. We then explained how data fusion algorithms can be used for estimating unknown parameters in a system, such as the path loss parameters for an RSS-based positioning system. In the closing section of the current chapter, some alternative approaches were presented, the first using a fingerprinting technique and the second using a time series approach.

6

Fundamentals of Tracking

6.1 Introduction

In the context of wireless positioning, one of the typical problems is to track a moving MS with time. The simplest approach would be to completely disregard any structural information contained in the sequence of observations and continuously run the estimation algorithms independently every time new observations were available. As was shown in Chapter 5, there are several tools that consider time history by propagating past knowledge into the present and future. As was previously seen, this approach provides the foundations for tracking moving devices.

This chapter starts with a conceptual explanation of the fundamental differences and the increment in complexity related to the movement of MSs. Then, this explanation is extended to the introduction of cooperative approaches. While taking movement into account constrains the position estimators by forcing a propagation of information along time, cooperative positioning influences the position estimators by constraining the estimation of the position of an MS to the position of neighboring MSs.

In order to guide the reader through the field of tracking applications, the present chapter will also introduce several of the most common mobility models. The chapter explains conventional models, geographically-constrained models, group mobility and cooperative mobility. These models are commonly the initial step for any further analysis concerning tracking algorithms.

As a generalized case of the tracking of wireless devices, the first step concerns the treatment of the raw measurements. In this chapter, a simple mitigation algorithm is explained as a solution for excluding outliers in the observations. More complex models are introduced in Chapter 7. Then, the theory behind tracking both nonmaneuvering and maneuvering MSs is explained. Finally, as the last step of a tracking application, it is possible to have a methodology for learning about typical

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

user movements. This information can, for instance, be used to analyze further activities of the MSs.

6.2 Impact of User Mobility on Positioning

This section highlights the important concepts that arise when mobility and cooperation are introduced into positioning applications. As it is conceptually explained below, these concepts increase the complexity of the problem.

6.2.1 Localizing static devices

Positioning static devices is, conceptually, the simplest scenario. Since the device is static, the mobility model does not depend on the time component. This fact constrains the problem of positioning a wireless device to the problem of finding its time-independent spatial coordinates. It is commonly possible to determine these coordinates based on indirect observations, i.e., physical phenomena that can be measured and have a relation to those coordinates (see Chapter 4). Mathematically, this means:

$$Z = \mathcal{H}(X), \tag{6.1}$$

where Z is the observable physical phenomena, X is the time-independent position coordinates and $\mathcal{H}(\bullet)$ is a function that relates the two variables. From Equation (6.1), it is possible to see that given two observations Z and two independent linear functions $\mathcal{H}(\bullet)$, the two coordinates (or three, depending on whether 2D or 3D positioning is being performed) can be calculated from:

$$X = \mathcal{H}^{-1}(Z). \tag{6.2}$$

It is important to note, however, that if $\mathcal{H}(\bullet)$ is nonlinear, two independent functions may not suffice.

6.2.2 Added complexity in tracking

When we assume movement of the wireless devices, the problem gets considerably more complex. Contrarily to the case in Section 6.2.1, the derivatives of the coordinates with respect to time are not zero, meaning that it is necessary to consider the derivatives of the space coordinates in Equation (6.1), such as

$$Z = \mathcal{H}(X, \dot{X}), \tag{6.3}$$

which implies that at least twice the number of measurements is necessary to determine the position coordinates and their derivatives compared with the static case. Note that Equation (6.3) could also be augmented with higher-order derivatives; however, due to the noise existing in typical wireless positioning systems, only the first derivative is considered. Higher orders are commonly difficult to determine.

6.2.3 Additional knowledge in cooperative environments

An advanced and more recent concept in wireless positioning is the use of cooperative schemes among several users. These schemes can equally well be used in the positioning of either static or moving devices. To simplify this concept though, this section assumes static devices, since cooperative positioning of moving devices may introduce problems due to the proximity among the terminals.

Cooperative schemes assume that wireless devices are able to obtain measurements of some kind among them, i.e., they are able to relatively position themselves. A typical example is short-range communications in ad hoc mode. Cooperation permits to augment the system of equations in Equation (6.1) for example in a system with two users, we have:

$$\begin{cases} Z_1 = \mathcal{H}_1(X_1), \\ Z_2 = \mathcal{H}_2(X_2), \\ Z_c = \mathcal{H}_c(X_1, X_2), \end{cases} \tag{6.4}$$

where the indices 1 and 2 identify the two wireless devices and the index c identifies the cooperative *part*. As Equation (6.4) shows, the cooperative factor augments the range of physical phenomena used for positioning wireless devices compared with the individualistic approach of Section 6.2.1.

One of the advantages of cooperative schemes is that by correlating the positions of the wireless devices, it is possible to minimize propagation effects such as shadowing. As Chapter 3 has explained, the shadowing effect can introduce bias into the positioning estimators. Since this bias is dependent on the position of the user, correlating the users tends to minimize the effect of this bias.

6.3 Mobility Models

A particulary important aspect of the tracking of moving devices concerns mobility models. By definition, a mobility model is a mathematical or conceptual formulation of rules that dictate and resemble the movement of the MSs depending on the scenario considered. In this section, several models for individual mobility, group mobility and social based mobility are described.

6.3.1 Conventional models

Conventional models are models that present a deterministic behavior. These models are commonly given by the basic laws of physics, namely Newton's laws of motion and equations of motion derived from them, available widely in the literature. For instance, the equations:

$$\dot{r} = \dot{r}_0 + \ddot{r}t \quad \text{and} \quad r = r_0 + \dot{r}_0 t + \frac{1}{2}\ddot{r}t^2 \tag{6.5}$$

represent uniformly accelerated linear motion. In Equation (6.5), r represents the movement in a certain direction, and t the time elapsed since time t_0, the initial state of the movement.

6.3.2 Models based on stochastic processes

6.3.2.1 Brownian-motion model

The Brownian-motion model is a random-movement model that describes a randomized movement of particles in a certain space. This model is often used in areas as diverse as the mechanics of fluids and finance, and in positioning applications. The model is simple and it is often used because of its mathematical convenience rather than its accuracy.

Consider the example of a million marbles inside a box where they all touch the bottom of the box in such a way that there is still plenty of space for marbles to move. Assume now that the box starts shaking and that all marbles start moving randomly due to the movement of the box and clashes with neighboring marbles. If we focus on a single marble, it is possible to see that it describes a randomized path. This path can be seen as the result of a Brownian motion.

The process that dictates a Brownian motion is equivalent to a Wiener process. A Wiener process is a continuous-time stochastic process that can be considered as the *limit* scenario of a random walk (see Section 6.3.2.2). A Wiener process W_t is characterized by three facts:

- $W_0 = 0$.
- W_t is almost surely[1] continuous.
- W_t has independent increments with distribution $W_t - W_s \sim \text{Norm}(0, t - s)$, for $0 \leq s < t$.

The condition that it has independent increments means that if $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2$ then $W_{t_1} - W_{s_1}$ and $W_{t_2} - W_{s_2}$ are independent random variables, and a similar condition holds for n increments.

A Wiener process has the following properties:

$$f_{W_t - W_s}(x) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-x^2/(2(t-s))}, \tag{6.6}$$

$$\text{E}[W_t - W_s] = 0, \tag{6.7}$$

$$\text{E}[(W_t - W_s)^2] - \text{E}^2[W_t - W_s] = t - s, \tag{6.8}$$

$$\text{cov}(W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}) = \min(t_1 - s_1, t_2 - s_2) \tag{6.9}$$

and

$$\text{corr}(W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}) = \frac{\min(t_1 - s_1, t_2 - s_2)}{\sqrt{(t_1 - s_1)(t_2 - s_2)}}. \tag{6.10}$$

[1] We say "almost surely" when a certain event has probability close to 1 to happen.

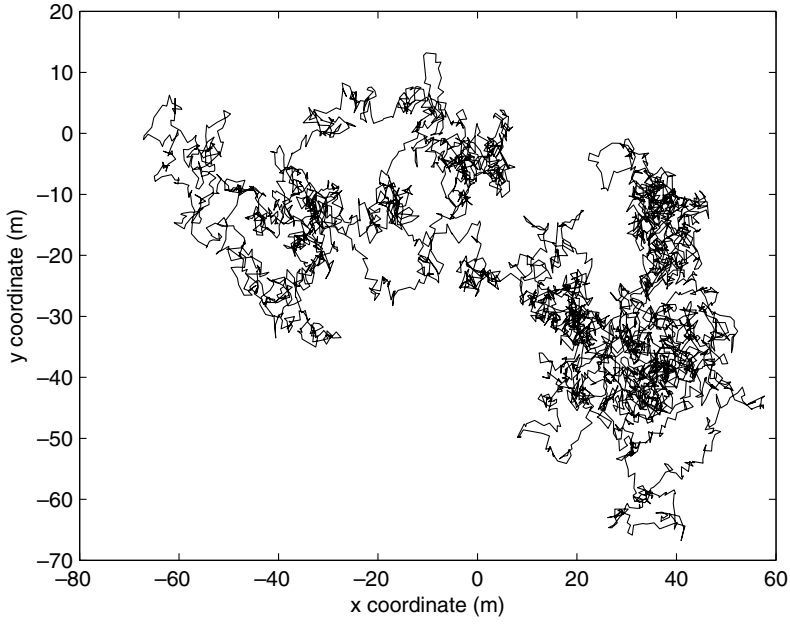


Figure 6.1 Example of a Brownian motion.

In physical terms, Equation (6.7) says that whatever path the MS follows, the average position is always the starting point. Figure 6.1 illustrates an example of a Brownian-motion movement, which shows a chaotic behavior.

6.3.2.2 Random walk model

The random walk is a discrete motion model which evolves with time in constant steps. At each time step, the model defines the next movement by randomly choosing a direction in which the next constant step is taken. When the step size tends to zero, the random walk is equivalent to the Brownian motion. If the step size s tends to 0, one needs to walk L/s^2 steps in order to *approximate* a Brownian motion of length L. Depending on the context, a random walk is often considered as a Markov chain (see Section 6.3.2.5), where the model evolves in time according to state transitions defined in the chain. Other approaches of interest consider random walks on graphs, on lines, in planes, on maps, in higher-order dimensions, in groups or in sets.

The random walk can be defined as follows:

- $W_0 = 0$.
- W_k is discrete.
- W_k has independent increments of either $-s$ or s, each outcome with probability $1/2$, where s defines the step size.

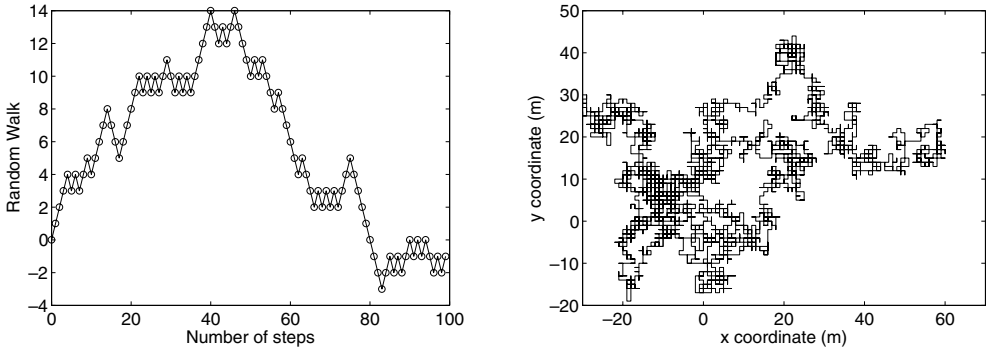


Figure 6.2 Random walk in one dimension (left) and two dimensions (right). Note that the circles that identify the steps in the left plot are not used in the right plot for easier reading of the figure.

A one-dimensional random walk can be defined as follows:

- $W_k = 0$ for $k = 0$.
- $W_k = \sum_{k=1}^{n} W_{k|k-1}$, where $W_{k|k-1}$ is either $-s$ or s, each outcome with probability $1/2$.

Figure 6.2 shows a random walk in one dimension (left plot) and in two dimensions (right plot), both with unit step size. The random walk presented in Figure 6.2 (left) is commonly called the simple random walk on $\mathbb{Z}$.

6.3.2.3 Waypoint random walk

The random waypoint model is a model often used as a benchmark in simulations of ad hoc networks when it is necessary to consider moving nodes. The model considers a predetermined area within which the nodes are allowed to move. A node, starting at a random position, selects the next destination, called a waypoint, within the simulation area based on a uniform distribution. Then, a sample velocity is drawn from a uniform distribution between a minimum and a maximum speed, and the node moves to the chosen destination at that speed. Optionally, the model can assume pause times when the node arrives at a waypoint. This pause time can be also drawn from a uniform distribution. Figure 6.3 exemplifies the random waypoint model. Although this model is used in a wide range of research work, one should be careful in using it. As detailed in the work of Yoon et al. (2003), this model fails to provide a steady state. This means that in certain simulations, where a steady state is necessary, such as in routing studies, the algorithm can lead to wrong conclusions. The problem is related to the randomness of the waypoints and the velocity at which the node travels until it gets to those waypoints. When waypoints are placed long distances apart and the velocity is relatively low, the time to get to a waypoint may be too large compared

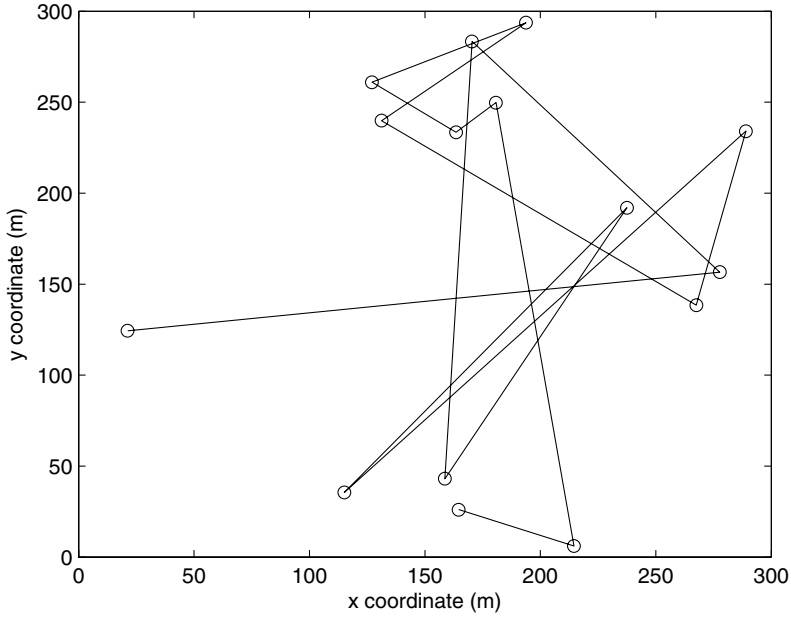


Figure 6.3 Example of a random waypoint movement.

with the simulation time. When the velocity tends to zero, the node takes an infinite time to arrive at the waypoint.

6.3.2.4 Gauss–Markov model

Another motion model that can appear in various formulations is the Gauss–Markov model. This model permits adaptation of the randomness of the mobility pattern by using a parameter γ that weights the random part of the model against the deterministic part. At fixed time steps, the movement is updated based on three parameters: (i) the previous position; (ii) the mean speed and direction; and (iii) a random variable. This updating is given by the following equations:

$$s_n = \gamma s_{n-1} + (1-\gamma)\bar{s} + \sqrt{(1-\gamma^2)}\, s_{x_{n-1}} \tag{6.11}$$

and

$$d_n = \gamma d_{n-1} + (1-\gamma)\bar{d} + \sqrt{(1-\gamma^2)}\, d_{x_{n-1}}, \tag{6.12}$$

where $0 \leq \gamma \leq 1$ is the tuning factor, s_n and d_n are the velocity and direction at time t_n, $\bar{s}$ and $\bar{d}$ are the mean velocity and direction, and $s_{x_{n-1}}$ and $d_{x_{n-1}}$ are random variables from a Gaussian distribution. Given Equations (6.11) and (6.12), it is possible to obtain the Cartesian coordinates as:

$$x_n = x_{n-1} + s_{n-1}\cos(d_{n-1}), \tag{6.13}$$

$$y_n = y_{n-1} + s_{n-1}\sin(d_{n-1}), \tag{6.14}$$

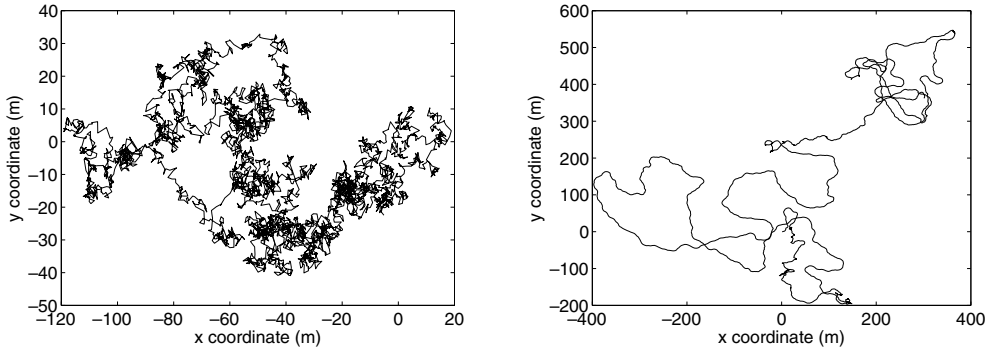


Figure 6.4 Example of a Gauss–Markov movement with $\gamma = 0.1$ (left) and $\gamma = 0.99$ (right).

where (x_n, y_n) are the Cartesian coordinates x and y where the mobile device is placed at time t_n. By inspecting the above equations it is possible to see that when γ tends to 0, the Gauss–Markov model approximates the Brownian-motion model. On the other hand, when γ tends to 1, the model approximates a linear movement.

Figure 6.4 shows two different examples of the Gauss–Markov model with two different values of γ. By comparing the two plots in Figure 6.4, it is possible to see that the higher the parameter γ, the higher is the tendency to maintain the direction of movement. Thus, by dynamically varying the parameter γ depending on the position obtained from the model, it is possible to differentiate between, for instance, areas with great flexibility of movement and areas with low flexibility of movement. For instance, in indoor scenarios, one may want movements in rooms to be more flexible with respect to changes in direction, while in corridors one may want movements to be more constrained.

6.3.2.5 Models based on Markov chains

Markov chains are stochastic processes that comply with the Markov property, i.e., future states depend only on the present state and a fixed number m of past states, and not on states from the more distant past. The Markov chain in this case is said to be of order m:

$$\mathrm{P}(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) = \mathrm{P}(X_n|X_{n-1}, X_{n-2}, \ldots, X_{n-m}). \tag{6.15}$$

In practice, a Markov chain is a system represented by a set of possible states with different transition probabilities between them. The system can either evolve by changing state or stay in the same state, depending on the probabilities associated with the state transitions. The random walk described in Section 6.3.2.2 is an example of a Markov chain.

In this section we show another example, obtained from Chiang (1998). The movement model is a simple Markov chain where the Cartesian coordinates are

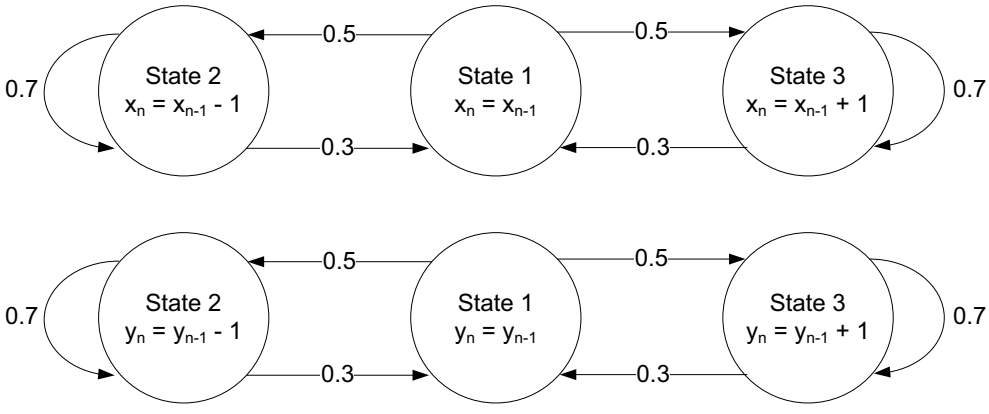


Figure 6.5 Example of a Markov chain motion model. The model assumes that the movements in each coordinate are independent from each other and that the MS can be in one of three states: static, decreasing by one unit or increasing by one unit.

treated independently and, for each coordinate, there are only three possible states: the coordinate value decreases by one unit, maintains its value or increases by one unit. This chain, shown in Figure 6.5, has its states represented by circles, while possible transitions between states are represented by directed arrows. The transition probabilities between states are given by the values associated with the arrows. Using this chain, it is possible to sample different movement paths. Figure 6.6 shows a possible sampling of a movement path using the Markov chain shown in Figure 6.5. As it is possible to see, this model favors linear movements rather than the chaotic movement that is characterized by Brownian motion. This linear-movement behavior may resemble the movement of people, as people tend to walk in segmentwise linear trajectories.

Another example of a Markov chain is shown in Figure 6.7, where only two states are possible. This is a simple model that could surely be implemented in several ways; however, it is still a simple and good example of the use of Markov chains. The model assumes a constant velocity of movement and the direction of movement θ is being governed by a Markov chain. Inspecting Figure 6.7, it is possible to see that once the process is in state 1, it has a 90% probability of maintaining the same direction and a 10% probability of changing direction. In order for the direction to change, the system must be in state 2, where a Gaussian-distributed random variable w_θ is added to the current direction. Once in this state, the system evolves towards state 1 with 100% probability. Figure 6.8 shows one sampling of a possible path obtained from this model. As in the case of the trajectory in Figure 6.6, this model also resembles the segmentwise linear trajectories followed by people.

Markov chains have great flexibility for modeling a wide variety of movement patterns. For instance, if the states represent actual positions in space, the chain can be designed in order to model a geographical-restriction model.

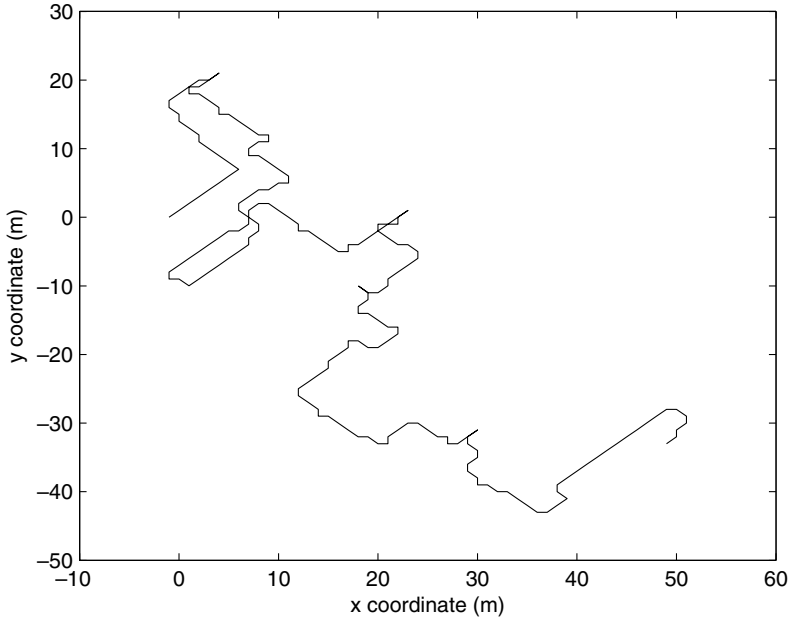


Figure 6.6 Example path sampled from the Markov model shown in Figure 6.5.

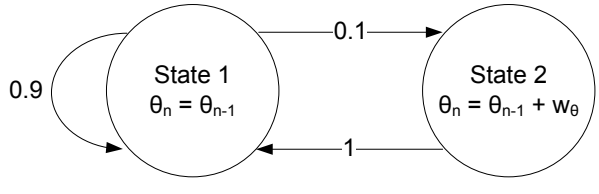


Figure 6.7 Example of a Markov chain motion model. The velocity is assumed constant and the direction to be variable. Each change in direction in state 2, follows a Gaussian distribution.

6.3.3 Geographical-restriction models

Geographical-restriction models are models that constrain the movement to a certain portion of space, to discrete trajectories or to discrete sets of positions. Some algorithms for modeling movement according to map constraints are good examples of geographical-restriction models.

6.3.3.1 Pathway mobility model

One example of the pathway mobility model that is widely present in the literature is the Manhattan mobility model proposed by the 3GPP (ETSI 1998). The Manhattan

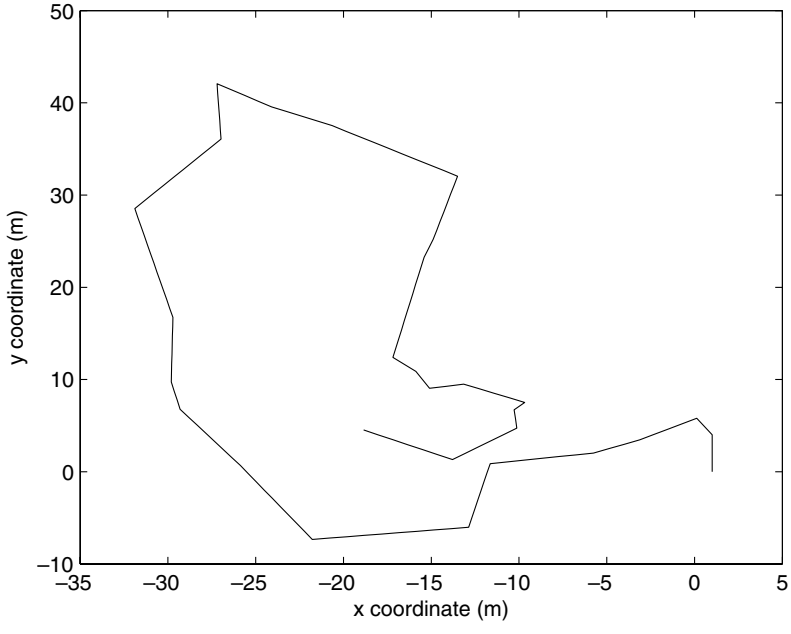


Figure 6.8 Example path sampled from the Markov model shown in Figure 6.7.

model has its widest application in the field of intervehicle communications. The model assumes a map with the typical shape of the streets of Manhattan. This urban map, pictured in Figure 6.9, permits users to move along the streets with a predefined probability of changing direction at every crossing. At every crossing, the probability of continuing in the same direction is 0.5, while the probability of turning either left or right is 0.25. Turning back has zero probability. Although the 3GPP proposed a model with defined speed limits and well-defined discretization, more generalized models can be derived. Figure 6.10 shows a single run of the Manhattan model.

Another example is the freeway model (Bai et al. 2003). This model emulates the mobility behavior of vehicles on a freeway. The model assumes that there are several freeways, where each freeway has several lanes in each direction. The mobility is restricted to the paths of the freeways with variable movements. The model assumes that several nodes move on a freeway and defines a safety distance between nodes, i.e., the mobility model must take into account neighboring nodes and avoid exceeding the velocity of the *preceding* node.

6.3.4 Group mobility models

In contrast to the previous models, where nodes move independently, group mobility models assume that there is a correlation among the various wireless devices. This may be the case when patrol, rescue or military teams are exploring an area. The movement of all the individuals is commonly strategic and is correlated in the sense

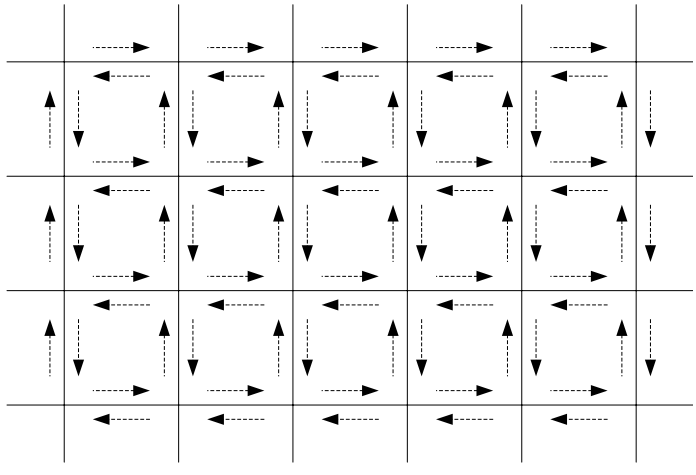


Figure 6.9 Example of map used by the Manhattan model.

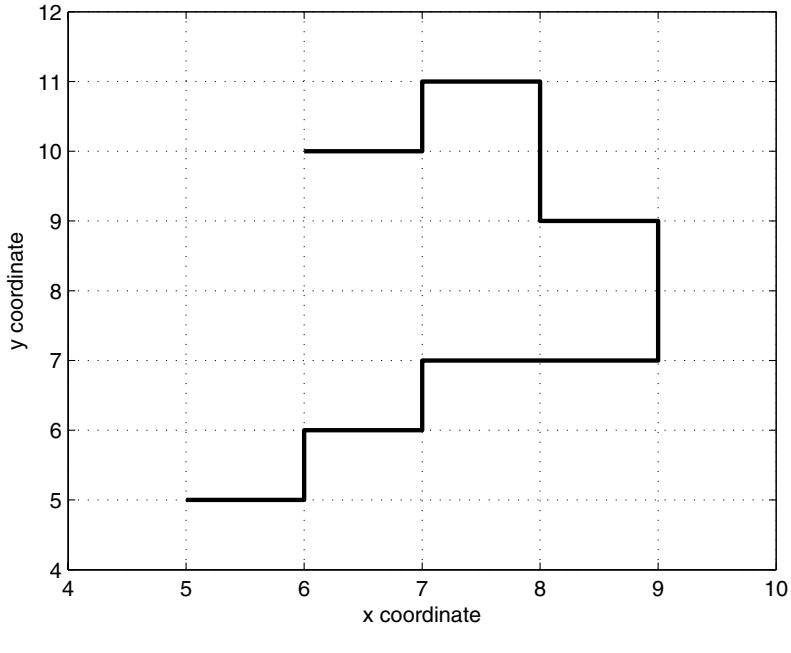


Figure 6.10 Single run of the Manhattan model. At every crossing, the probability of continuing straight is 0.5, of turning either left or right is 0.25, and of turning backwards is 0.

that everyone has a common goal. In practice, group models constrain the trajectories of each individual by establishing correlation properties among them.

6.3.4.1 Reference point group mobility model

The reference point group mobility model is a model that provides a foundation for several other derived models. This model represents the movement of the center of the group in addition to the movement of the individuals within the group. Thus, it is convenient to assume that the group moves according to a certain mobility model with predetermined parameters, while the individuals move within the group according to a second mobility model with another set of parameters. For instance, assuming that both the group and the individuals move according to a waypoint random model, the position of each individual i at a certain time t is determined by:

$$X_i(t) = X_g(t) + r_i, \quad X_g(t) = r_g, \tag{6.16}$$

where r_i is a random variable with a zero mean Gaussian distribution and variance σ_i, and $X_g(t) = r_{\mathrm{g}}$, with r_{g} being a second random variable with a uniform distribution across the entire scenario. In practice, Equation (6.16) represents the vectorial sum of two random variables, one dictating the movement of the group and the other dictating the movement of each individual within the group.

Figure 6.11 shows a run of the reference point group mobility model when both the group and each individual move according to a random waypoint model. As it is possible to see, there is a strong correlation among the trajectories of the nodes, which is a consequence of Equation (6.16).

In a more general case, this model can be seen as a framework where the mobility of a group and of individuals within the group are treated separately and then combined in a vectorial summation. Given this generalization, it is possible to assume any model for the group and for the individuals. For instance, it is possible to assume a model such as people moving in a bus, i.e., the relative movement of the individuals within the group is zero and only the group moves. At the other extreme, it is possible to assume that the group has no movement and the individuals move according to some given model, resulting in a type of model where individuals move independently.

6.3.4.2 Correlation group mobility model

Another type of mobility model considers group mobility as a correlation among several devices. One device is taken as a reference and all the others are bound to the trajectory of this reference according to a certain correlation factor. The correlation factor can be seen as a variable that defines the affinity of a particular individual to the entire group. Thus, assuming that $X_1(t) = r_1$ (where r_1 is a random variable with a uniform distribution, taking into account the entire simulation area), the positions $X_i(t)$ of the other individuals are defined as:

$$X_i(t) = \rho_i X_1(t) + \sqrt{1 - \rho_i^2} r_i, \tag{6.17}$$

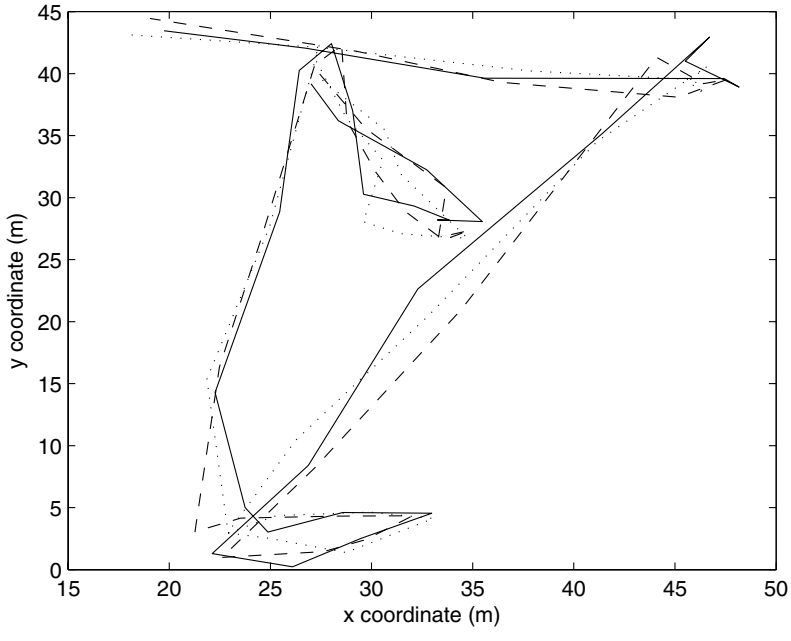


Figure 6.11 Sample pattern obtained by a simulation of a reference point group mobility model. Each line represents the trajectory of a different individual.

where ρ_i is the correlation factor between the individual i and the reference, and r_i is a random variable with a uniform distribution. In general, r_i is taken as a generalized distribution on the entire simulation area and is the same for every device. In contrast, the correlation factor ρ_i can be considered individual-dependent. This allows one to define different affinities between each device and the entire group. Figure 6.12 shows two different runs of this group mobility model. In the left plot, where the correlation factor is 0, it is possible to see that the trajectories do not look alike among the different individuals. In fact, this is an extreme case, where nodes have independent trajectories and thus the model results in individual mobility instead of a group mobility. In the right plot, with a correlation factor of 0.999, it is possible to see that the trajectories of the individuals have structural similarities, where individuals do move as a group.

6.3.5 Social-based models

Social-based mobility models started to be investigated during the boom in social networking a few years ago. In contrast to group mobility models, which simulate the movement of individuals both within a group and as a group, social-based models have an extra dimension which takes into account the social behavior of individuals and groups with their peers. As wireless devices to be localized or tracked are

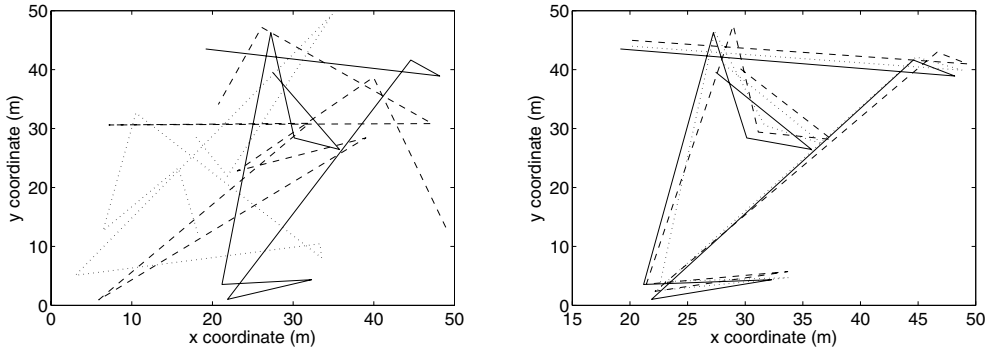


Figure 6.12 Example of the correlation group mobility model with two different values for the correlation factor. In the left plot the correlation factor is 0, while in the right plot the correlation factor is 0.999.

commonly carried by people, social models try to integrate people's social behavior into their movement.

6.3.5.1 Model based on a sociability factor

This model, proposed by Musolesi et al. (2004), is based on statistical generation of social relationships. Social behavior is represented by a weighted graph, where nodes are wireless devices and edges represent the social affinity between the persons carrying those wireless devices. This social affinity between the nodes is defined by an interaction indicator m, which is then randomly chosen between 0 and 1, where 0 indicates no interaction and 1 represents strong interaction. The generation of the values of m can be done using a uniform distribution or any other distribution that is intended to simulate the desired society.

The interaction matrix M representing the social graph is then defined as:

$$M = \begin{bmatrix} 1 & m_{1,2} & \cdots & m_{1,n} \\ m_{2,1} & 1 & \cdots & m_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & 1 \end{bmatrix}, \tag{6.18}$$

where $m_{i,j}$ represents the interaction indicator between nodes i and j. The diagonal elements are conventionally set to 1 and the matrix is assumed to be symmetric, i.e., the interaction between the two nodes of a pair is reciprocal.

Given the interaction matrix of Equation (6.18), the social factor S of each individual is calculated as:

$$S = \frac{\sum_{j=1, j\neq i, m_{i,j}>\xi}^{n} m_{i,j}}{v}, \tag{6.19}$$

where ξ is a connection threshold below which individuals are assumed to be socially disconnected, and v is the total number of individuals, so that $m_{i,j} > \xi$. Once the social factor is known, the nodes are sorted inversely according to this factor. Then, the first node is placed in a random group. The following nodes i are placed in a randomly chosen group that has no nodes. When each group has one node, the remaining nodes are placed into the groups according to their group attraction, which is calculated as:

$$A_{i,G} = \frac{\sum_{j=1, j \in G}^{n} m_{i,j}}{w}, \tag{6.20}$$

where G stands for group and w is the number of members in the group. As an alternative, Equation (6.20) was rewritten in Musolesi et al. (2004) in order to accommodate a relation between the attraction and the distance between the node and the group to which attraction is being calculated. Optionally, a predetermined number of nodes with the lowest group attraction can be placed randomly within the entire simulation area instead of assigning them to a group.

Once the placement of all the nodes is accomplished according to the aforementioned social rules, the movement is modeled as a discrete-event simulation. Then, any movement is simulated according to a social variant of the waypoint random walk model described in Section 6.3.2.3. The major difference is the introduction of decision points originating from social behavior. First, an entire group is assumed to be moving, then the movement of each individual is augmented by movement within the group and, finally, nodes outside the groups move within the entire simulation area. Once the nodes arrive at a goal given by the waypoint random walk model, social decisions are made based on a uniform distribution, i.e., a node decides to move outside of any group if a sampling of the uniform distribution is lower than the social factor of that node. Otherwise, the node stays within a group, but not necessarily the group that it belonged to until then. The node will move to the group to which it has the biggest group attraction. If that is the same group, the node selects the next goal within that same group. Otherwise, it moves to the second group.

Figure 6.13 shows a single run of the social mobility model. The left plot shows the placement of the groups and the placement of five nodes within the simulation area. The two squares define the limits of each group. In the right plot of Figure 6.13, it is possible to see the outcome of two simulation steps. In this simulation, it was assumed that the goals set by each node were accomplished within a single simulation step. As an alternative, one could set a constant velocity and then assume that goals are reachable only when the entire distance required to get to that specific goal has been covered. This approach specifies that new goals are sampled only when the target gets to the previously defined goal. If in any simulation step the node is moving toward a certain goal, that movement must be respected.

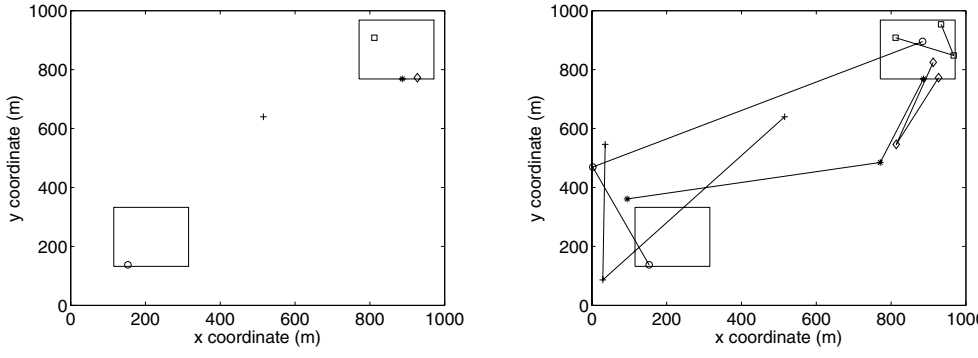


Figure 6.13 Simulation of the sociability factor model. The left plot shows the initial random placement of the nodes, and the right plot shows a simulation of a movement where nodes move two steps.

6.4 Tracking Moving Devices

In both positioning or tracking applications, one of the first phases concerns the treatment of the measurements. This step regards the initial analysis that is necessary in order to identify some effects that can be canceled at a very early stage. For instance, detecting outliers and ignoring them is a type of analysis that can be done before the actual position estimation algorithm. Although Chapter 7 introduces some of these techniques, a technique for detecting outliers will be described in this section.

Regarding the positioning mechanism, the simplest assumption in the case of tracking moving MSs is that the kinetic laws governing the movement do not change with time, but that the position coordinates and their derivatives do not change with time. Instead, it means that the set of equations governing the movement are the same along the entire trajectory. This type of movement can be defined as a nonmaneuverable movement.

A natural step ahead of tracking a moving MS is to detect maneuvers in the movement of the MS. As it has been seen in previous chapters, the major problem is that the data available for estimating the position is typically corrupted by noise components with high variances compared with the magnitude of the expected values of the measurements. Moreover, the precise distributions of these noise components are unknown. Thus, it is common to approximate the noise distributions by known distributions. This approach results in estimation algorithms that cannot, in a simple fashion, track MSs with frequent maneuvers. The term “frequent” is defined here, loosely and fuzzily, as the inverse of the time that an algorithm needs to “understand” that changes have occurred in the MS movement rules. If these maneuvers can be detected, the common strategy is to analyze the sequence of observations and detect systematic changes in the expected value so that it is possible to infer changes in movement laws.

On top of these solutions for wireless positioning, it is possible to consider mechanisms that are able to learn further information from the estimated positions or trajectories, for example clustering of MSs or trajectory-learning methods. These type of algorithms, as will be seen in Chapter 9, provide a foundation for more complex cooperative schemes. This section shows in addition some of the algorithms used for postprocessing the results of the positioning or tracking algorithms.

6.4.1 Mitigating obstructions in the propagation conditions

Obstructions frequently occur in wireless communications. These obstructions can be found by statistical treatment of the measurements. For example, an obstruction in the transmitted signal can be observed at the receiver through a higher delay or a stronger attenuation. Thus, this type of effect can be detected, compensated or possibly ignored. In mathematical terms, these measurements are considered as outliers.

An outlier is detected when a "large" innovation happens, which can be discovered by using the normalized innovations squared:

$$\varepsilon_k = \tilde{Z}_k^{\mathrm{T}} S_k^{-1} \tilde{Z}_k, \tag{6.21}$$

where $\tilde{Z}_k$ and S_k are given by Equations (5.78) and (5.79), respectively.

Since $\tilde{Z}_k$ is by assumption Gaussian distributed, it is possible to prove that Equation (6.21) is a chi-squared random variable with l degrees of freedom (where l is the dimension of the observation vector Z).

Based on the assumption that ε_k is chi-squared, a possible solution for detecting an outlier would be to check when ε_k exceeds a predetermined threshold. Thus, by ignoring the measurements that exceed this predetermined threshold, obstructions can be handled, reducing the effect on the estimated positioning information.

6.4.2 Tracking nonmaneuvering targets

Positioning nonmaneuverable MSs in wireless communications is typically accomplished by assuming a kinetic law given by the following linear differential equation:

$$\dot{X}(t) = FX(t) + G(t), \tag{6.22}$$

where

$$X(t) = \begin{bmatrix} x(t) \\ y(t) \\ \dot{x}(t) \\ \dot{y}(t) \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad G(t) = \begin{bmatrix} 0 \\ 0 \\ g_x(t) \\ g_y(t) \end{bmatrix}, \tag{6.23}$$

and where $g_x(t)$ and $g_y(t)$ are independent zero-mean Gaussian distributions with standard deviations σ_x and σ_y, respectively. The covariance matrix of $G(t)$ is

defined as:

$$\mathrm{E}[G(\tau)G(\upsilon)^{\mathrm{T}}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_x\delta(\tau-\upsilon) & 0 \\ 0 & 0 & 0 & \sigma_y\delta(\tau-\upsilon) \end{bmatrix}, \tag{6.24}$$

where $\delta()$ is the Dirac delta function. Note that these equations are restricted to the first derivative of the position, i.e., the velocity. The problem can be easily augmented for the case where the acceleration is also of interest. A consequence of this augmentation is that the problem becomes more complex and the position estimators may diverge quicker than in the case where the highest derivative of position considered is the velocity. In addition, it is worth mentioning that another problem is related to the fact that major wireless technologies obtain channel measurements with a frequency that may not be as large as necessary to have enough samples to estimate changes in the acceleration. For these reasons, the model is kept simple in the context of wireless positioning. Note as well that Equations (6.22) and (6.23) represent a straight-line movement, although several other approaches could be considered, which could be derived in an equivalent manner.

In order to obtain a recursive expression for the movement, it is necessary to solve Equation (6.22). The first step is to multiply each side of Equation (6.22) by e^{-Ft}:

$$e^{-Ft}\dot{X}(t) = e^{-Ft}FX(t) + e^{-Ft}G(t), \tag{6.25}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{-Ft}X(t) = e^{-Ft}G(t). \tag{6.26}$$

Then, integrating both sides over the interval $[t,\ t+\mathsf{t}]$, we get:

$$e^{-F(t+\mathsf{t})}X(t+\mathsf{t}) - e^{-Ft}X(t) = \int_t^{t+\mathsf{t}} e^{-F\tau}G(\tau)\,\mathrm{d}\tau, \tag{6.27}$$

$$X(t+\mathsf{t}) = e^{F\mathsf{t}}X(t) + \int_t^{t+\mathsf{t}} e^{F(t+\mathsf{t}-\tau)}G(\tau)\,\mathrm{d}\tau. \tag{6.28}$$

By evaluating the matrix exponential $e^{F\mathsf{t}}$ by series expansion, it is possible to see that

$$e^{F\mathsf{t}} = I + \mathsf{t}F + \sum_{i=2}^{\infty}\frac{\mathsf{t}^i F^i}{i!}, \tag{6.29}$$

where the last term equals the zero matrix, since F to a power larger than 1 also equals the zero matrix,

$$A \equiv e^{F\mathsf{t}} = I + \mathsf{t}F = \begin{bmatrix} 1 & 0 & \mathsf{t} & 0 \\ 0 & 1 & 0 & \mathsf{t} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.30}$$

Thus, the expected value of Equation (6.28) is given by:

$$\mathrm{E}[X(t+\mathsf{t})] = A\mathrm{E}[X(t)], \tag{6.31}$$

with A given by Equation (6.30).

Regarding the covariance matrix of Equation (6.28), it is possible to see that

$$Q \equiv \mathrm{E}[X(t+\mathsf{t})X(t+\mathsf{t})^{\mathrm{T}}] \tag{6.32}$$

$$= E\left[\left(\int_t^{t+\mathsf{t}} e^{F(t+\mathsf{t}-\tau)}G(\tau)\,\mathrm{d}\tau\right)\left(\int_t^{t+\mathsf{t}} e^{F(t+\mathsf{t}-\upsilon)}G(\upsilon)\,\mathrm{d}\upsilon\right)^{\mathrm{T}}\right] \tag{6.33}$$

$$= \int_t^{t+\mathsf{t}}\int_t^{t+\mathsf{t}} e^{F(t+\mathsf{t}-\tau)}E[G(\tau)G^{\mathrm{T}}(\upsilon)]e^{F(t+\mathsf{t}-\upsilon)^{\mathrm{T}}}\,\mathrm{d}\tau\,\mathrm{d}\upsilon \tag{6.34}$$

$$= \int_t^{t+\mathsf{t}} e^{F(t+\mathsf{t}-\tau)}\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&\sigma_x^2&0\\0&0&0&\sigma_y^2\end{bmatrix}e^{F(t+\mathsf{t}-\tau)^{\mathrm{T}}}\,\mathrm{d}\tau, \tag{6.35}$$

where the step from Equation (6.34) to Equation (6.35) is a consequence of the Dirac delta function included in Equation (6.24), which is forced to be nonzero only when $\tau = \upsilon$. The solution for Equation (6.35) can be determined, after manipulation, to be:

$$Q = \mathsf{t}\begin{bmatrix}\sigma_x^2\mathsf{t}^2/3 & 0 & \sigma_x^2\mathsf{t}/2 & 0\\ 0 & \sigma_y^2\mathsf{t}^2/3 & 0 & \sigma_y^2\mathsf{t}/2\\ \sigma_x^2\mathsf{t}/2 & 0 & \sigma_x^2 & 0\\ 0 & \sigma_y^2\mathsf{t}/2 & 0 & \sigma_y^2\end{bmatrix}. \tag{6.36}$$

Thereby, the typical movement laws used in estimating the positions of MSs in wireless technologies are given by the expected value of X, as in Equation (6.31), and the covariance matrix of the noise component, as in Equation (6.36).

6.4.3 Tracking maneuvering targets

This section introduces some of the algorithms used for detecting maneuvers. These algorithms can be considered as an add-on to the approach previously presented for tracking nonmaneuverable targets. The concept is that when a maneuver is detected, the positioning system takes action in order to recalculate the new movement parameters. This action can, for instance, momentarily increase the covariance of the noise component in the positioning estimators, i.e., increase the uncertainty in the position estimators.

6.4.3.1 Process adaptation using maneuver detection

This type of maneuver detection is closely related to the formulation of the Kalman filter. The strategy is to evaluate the innovation process as a function of time given by

Equation (5.78). As mentioned in Bar-Shalom et al. (2001), one maneuver detection scheme suitable for use when the noise components are Gaussian is the "white noise model with adjustable level".

A maneuver is detected when a sequence of "large" innovations happens, which can be discovered by using the normalized innovations squared:

$$\varepsilon_k = \tilde{Z}_k^{\mathrm{T}} S_k^{-1} \tilde{Z}_k, \tag{6.37}$$

where $\tilde{Z}_k$ and S_k are given by Equations (5.78) and (5.79), respectively.

Since $\tilde{Z}_k$ is by assumption Gaussian-distributed, it is possible to prove that Equation (6.37) is a chi-squared random variable with l degrees of freedom (where l is the dimension of the observation vector Z). Based on the assumption that ε_k is chi-squared, a possible solution for detecting a maneuver would be to check when ε_k exceeds a predetermined threshold. However, owing to the stochastic properties of ε_k, maneuvers can be incorrectly detected. To minimize this problem, an exponential discounted average can be used with a forgetting factor γ:

$$\varepsilon_k^{\gamma} = \gamma \varepsilon_{k-1}^{\gamma} + \varepsilon_k, \quad 0 < \gamma < 1, \quad \varepsilon_0^{\gamma} = 0, \tag{6.38}$$

where ε_k^{γ} is by the first-moment approximation chi-squared distributed with $l/(1-\gamma)$ degrees of freedom. If $l/(1-\gamma)$ is not an integer, the gamma function should be used. For additional details, see Bar-Shalom et al. (2001).

When a maneuver is estimated, i.e., ε_k^{γ} has exceeded the predefined threshold, the level of process noise is momentarily increased, meaning that the uncertainty in the estimation increases. Conversely, when ε_k^{γ} goes below the threshold, a lower process noise is used again.

6.4.3.2 Multiple-model approaches

An alternative approach to tracking maneuvering users is to use a bank of r different models, running in parallel. Then a decision is required on how they can be combined. The r models can be considered, for example, as modeling different movement dynamics of the MS, for example static vs. moving, forward movement vs. sideways movement or horizontal movement vs. vertical movement.

The simplest multiple-model approach assumes that the model that the system obeys is fixed, but unknown. Thus, the problem is to decide which one of the r models is the one that produces the output which is observed. This solution is suitable when the model does not change with time. The final estimation can be considered as a decision about which model is the most probable filter.

In a more advanced approach, the model that describes the system can undergo a switching process with time. These systems, also known as jump-linear systems or system-jumping processes, can be modeled by:

$$X_k = A(M_k) X_{k-1} + V(M_k) \zeta_k, \tag{6.39}$$

$$Z_k = H(M_k) X_k + W(M_k) \eta_k, \tag{6.40}$$

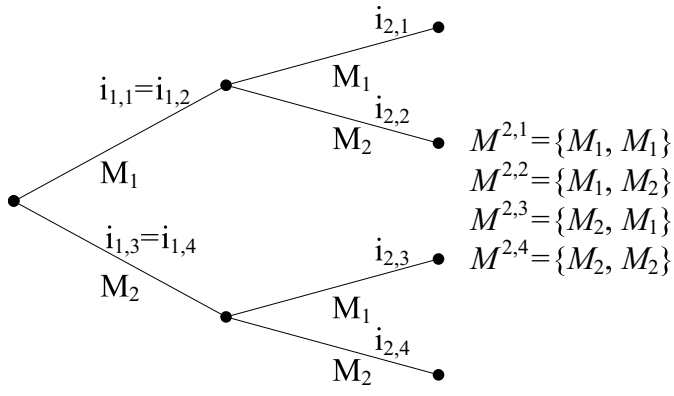


Figure 6.14 Example of execution of the multiple-model framework. Two filters run in parallel for two units of time. The result is four histories.

where M_k denotes the model at time k. The model that is active at time t_k is assumed to be one of the possible r models (or filters).

Assuming that the r filters are running in parallel, a decision about combining the filters can be taken by a switch variable Y_k. This variable specifies which movement pattern to use at each time instant t_k. In the case of Kalman filters, this switching variable decides which $\{A, Q, H, R\}$ matrices to use. The combination of the models is based on a weighted scheme where weights are given by $Pr(Y_k = i|Z_{1:k})$, i.e., the probability of using model i given measurements $Z_{1:k}$ from time t_1 until time t_k. This is called a "soft switching" scheme.

The lth model history (or sequence of models) through time k is denoted by:

$$M^{k,l} = \{M_{i_{1,l}}, \ldots M_{i_{k,l}}\}, \quad l = 1, r^k, \tag{6.41}$$

where $i_{\kappa,l}$ is the model index at time κ from history l, and

$$1 \leq i_{\kappa,l} \leq r, \quad \kappa = 1, \ldots, k. \tag{6.42}$$

Figure 6.14 shows a schematic representation of a multiple model running with two filters in parallel.

A fundamental problem of this multimodel framework is that running r filters in parallel represents a geometric complexity cost (see Figure 6.14). The reason is that at each time step, every single estimator output of the parallel filters must be input into all r filters.

For simplification and easier tractability of the mathematical formulation, it is common to assume that the mode switching has Markovian properties of the first order:

$$m_{ij} \equiv \mathrm{P}\{M(k) = M_j | M(k-1) = M_i\}. \tag{6.43}$$

The system of equations defined by Equations (6.39), (6.40) and (6.43) is a generalized hidden Markov model. Given this system, it is possible to obtain the

probability for each history (Bar-Shalom et al. 2001):

$$\mu^{k,l} = \frac{1}{c} \mathrm{P}\{Z_k | M^{k,l}, Z_{1:k-1}\} m_{ij} \mu^{k-1,s}, \tag{6.44}$$

where $i = s_{k-1}$ is the last model of the parent sequence s and c is the normalization constant. As it is shown in Equation (6.44), there is still the necessity to condition on the entire history, even though the switching model has Markovian properties.

Taking into account the entire history means manipulating an algorithm that grows exponentially in time. This behavior is clearly intractable after a few time steps, even for the smallest system with only two models. In order to solve this scalability problem, there are several approaches. The generalized pseudo-Bayesian (GPB) (Bar-Shalom et al. 2001) approach is an algorithm that keeps track of the recent history, collapsing the older part of the history into a single estimator. Thus, the first-order GPB approach collapses the history for time $k-1$ when the estimation is at time k. The second-order GPB approach collapses the history until time $k-2$. Correspondingly, the first- and second-order approaches use r and r^2 filters running in parallel. An alternative solution to the second-order GPB approach is the interacting-multiple-model approach, which reduces the number of parallel filters from r^2 to r (Murphy 1988).

6.4.4 Learning position and trajectory patterns

The learning of position patterns is related to artificial intelligence and machine learning. This is an enormous area with an extensive literature applied to a wide range of different applications. The reason why machine learning is such a wide area is related to the core problem that it addresses. In simple terms, the problem can be defined as: given a set of inputs and, optionally, the corresponding set of outputs, what is the function and/or parameters that relate the input data to the output data. If the output data is not available, machine learning attempts to do the same job based on predefined rules.

Based on the methodology for learning patterns, machine-learning algorithms can be classified into several groups. Most of the algorithms fall into one of the following groups:

- **Supervised learning** aims at estimating the relationship between the inputs and outputs of a system. By using the estimated relationship, these techniques are able to predict future outputs based on future inputs. As an example, one can think of the problem of determining the propagation parameters in a certain scenario, given a set of propagation measurements and the set of target positions that produced those measurements.

- **Unsupervised learning** aims at automatically discovering structures, groups, features or representations, given only the inputs in a system. An example might be the problem of clustering wireless devices in a network.

- **Reinforced learning,** given inputs, real-world actions and coefficients of reward or punishment aims at learning to select a sequence of actions resulting from future inputs in a way that maximizes the expected reward. A typical example is the type of learning used by robots.

6.4.4.1 The expectation maximization algorithm

The expectation maximization (EM) algorithm is an efficient iterative procedure to compute the maximum-likelihood estimator (MLE) in the presence of a hidden random variable. Each iteration of the algorithm consists of two steps: the expectation step (E-step) and the maximization step (M-step). The E-step consists in estimating the missing variables given the observed data and the current estimators for the model parameters. The M-step consists in maximizing the likelihood function under the assumption that the hidden variables are known.

Ignoring at this point the hidden variables, the problem is to find the parameters θ such that $\mathrm{P}(X|\theta)$ is maximum. In order to estimate θ, it is common to introduce the log-likelihood function, which is defined as:

$$L(\theta|X) = \ln \mathrm{P}(X|\theta). \tag{6.45}$$

The foundation of Equation (6.45) is that the parameter θ that maximizes $\mathrm{P}(X|\theta)$ also maximizes $L(\theta|X)$, owing to the monotonic property of the logarithm. The problem is that maximizing Equation (6.45) is typically intractable. The EM algorithm provides a framework which by including hidden variables makes the problem computationally simpler. These hidden variables can be either nonmeasurable variables of the system, as in the tracking application (see below), or purely a mechanism to ensure tractability of the problem:

$$L(\theta|X, W) = \ln \int_w \mathrm{P}(X, w|\theta), \tag{6.46}$$

where W is the hidden variable. Furthermore, it is possible to prove that maximizing Equation (6.46) is equivalent to iteratively maximizing (Borman 2004):

$$\mathrm{E}_{W|X,\theta}[L(\theta|X, W)]. \tag{6.47}$$

Based on Equation (6.47), is now possible to define the algorithm. The algorithm iteratively executes the following steps:

E-step: Determine the conditional expected value of Equation (6.47) given that θ is known:

$$\mathrm{E}_{W|X,\theta^{(t)}}[L(\theta|X, W)]. \tag{6.48}$$

M-step: Maximize Equation (6.47) with respect to θ:

$$\theta^{(t+1)} = \arg \max_{\theta^{(t)}} \{\mathrm{E}_{W|X,\theta^{(t)}}[L(\theta|X, W)]\}. \tag{6.49}$$

The initial value of $\theta^{(0)}$ is set manually as an expected value.

The EM algorithm permits several tracking applications, such as methods for learning user position and mobility. For instance, a simple example concerns clustering of position data. In the specific context of group mobility and cooperative positioning, clustering is an important matter. In a scenario where no prior clustering data is available, it is necessary to perform some kind of unsupervised learning in order to define a cluster. We assume that data points in clusters are formed according to a mixture of three Gaussian random variables, which are defined by:

$$\mathrm{f}(X|\theta) = \sum_{j=1}^{3} \frac{\tau_j}{(2\pi)|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(X-\mu_j)\right), \tag{6.50}$$

where $\theta = \{\mu_1, \mu_2, \mu_3, \Sigma_1, \Sigma_2, \Sigma_3, \tau_1, \tau_2, \tau_3\}$ is the parameter set, $\tau_j = \mathrm{P}(W = j)$, with W as the hidden variable that identifies each single Gaussian distribution, and μ_j and Σ_j are the mean and covariance of the Gaussian distribution j.

From Equation (6.50), it is possible to obtain the log-likelihood function, which is given by:

$$L(\theta|X, W) = \ln\left[\prod_{i=1}^{n}\sum_{j=1}^{3} \frac{\delta(W_i = j)\tau_j}{(2\pi)|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(X_i-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(X_i-\mu_j)\right)\right], \tag{6.51}$$

where $\{X_1, \ldots, X_n\}$ are n input data points for the random variable X and $\{W_1, \ldots, W_n\}$ represent realizations of the random variable W. Note that the latter can be seen as an artifact introduced in order to ease the algorithm, because these variables are not known from input data. By solving Equation (6.51), it is possible to obtain:

$$\begin{aligned} L(\theta|X, W) = \sum_{i=1}^{n}\sum_{j=1}^{3} \delta(W_i = j)\Bigg[&\ln(\tau_j) - \frac{1}{2}\ln|\Sigma_j| \\ &- \frac{1}{2}(X_i-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(X_i-\mu_j) - \ln(2\pi)\Bigg]. \end{aligned} \tag{6.52}$$

Note that in Equation (6.52) the ln() function is passed into the summation because $\delta(W_i = j)$ forces the summation in the variable j to have only one nonzero term. Applying Equation (6.52) in Equation (6.47), it is possible to obtain:

$$\begin{aligned} \mathrm{E}[L(\theta|X, W)] = \sum_{i=1}^{n}\sum_{j=1}^{3} \xi_{j,i}\Bigg[&\ln(\tau_j) - \frac{1}{2}\ln|\Sigma_j| \\ &- \frac{1}{2}(X_i-\mu_j)^{\mathrm{T}}\Sigma_j^{-1}(X_i-\mu_j) - \ln(2\pi)\Bigg], \end{aligned} \tag{6.53}$$

where Equation (6.53) corresponds to the E-step, and $\xi_{j,i}$ is given by:

$$\xi_{j,i} = \mathrm{P}(W|X, \theta)$$
$$= \frac{(\tau_j/((2\pi)|\Sigma_j|^{1/2})) \exp(-(1/2)(X_i - \mu_j)^{\mathrm{T}}\Sigma_j^{-1}(X_i - \mu_j))}{\sum_{k=1}^{3}(\tau_k/((2\pi)|\Sigma_k|^{1/2})) \exp(-1/2)(X_i - \mu_k)^{\mathrm{T}}\Sigma_k^{-1}(X_i - \mu_k))}. \quad (6.54)$$

At this point the M-step can be calculated by setting to zero the gradient of Equation (6.53) with respect to the variables θ and finding their zeros. The result is:

$$\tau_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\xi_{j,i}^{(t)}, \quad (6.55)$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n}\xi_{j,i}^{(t)}X_i}{\sum_{i=1}^{n}\xi_{j,i}^{(t)}}, \quad (6.56)$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{n}\xi_{j,i}^{(t)}(X_i - \mu_j^{(t)})^{\mathrm{T}}(X_i - \mu_j^{(t)})}{\sum_{i=1}^{n}\xi_{j,i}^{(t)}}. \quad (6.57)$$

As an example, assume that a possible sampling of the distribution of Equation (6.50) is the one shown in Figure 6.15 (left). This figure shows a hypothetical scenario of a room with three divisions. The data shows a hypothetical set of Cartesian points obtained by a positioning method. Suppose that the purpose is to group data points in order to infer further information, such as scenario mapping. Given these samples as inputs, the EM algorithm is run in order to cluster the data. The EM algorithm is composed by the iterative sequence of Equation (6.54) and Equations (6.55)–(6.57). The result of the EM algorithm for a mixture of Gaussian distributions is given in Fig. 6.15 (right). The initial values for the parameters were rather close to the real values. However, the higher the number of data points X_i, the more robust is the EM algorithm for handling the initial values properly. As the figure shows, the algorithm was able to clusterize the data given the sampling data. This is an unsupervised method for determining groups of data.

6.4.4.2 The *k*-means algorithm

The k-means algorithm defines a methodology to partition an entire set of elements into a predefined number of groups. Each element belongs to the group for which it has the lowest Euclidean distance to the average of all elements in that same group.

To formulate the algorithm, let us assume that a set of n elements $\{X_1, \ldots, X_n\}$ is to be partitioned into k groups $\{G_1, \ldots, G_k\}$ with $k < n$. The k-means algorithm *defines* that each element belongs to the group which satisfies the condition of minimum within-cluster sum of squares:

$$\hat{G} = \arg\min_{G}\sum_{i=1}^{k}\sum_{X_j \in G_i}\|X_j - \mu_i\|^2, \quad (6.58)$$

where μ_i is the mean value of the elements of G_i.

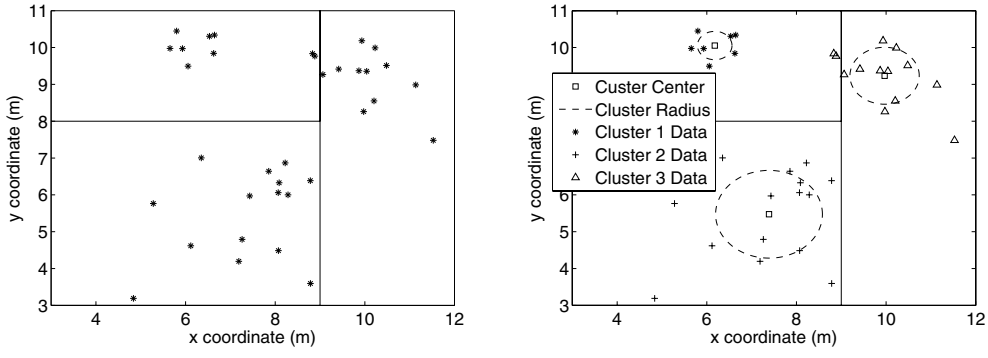


Figure 6.15 Clustering positioning data in order to identify different areas. The right plot shows the data points in a hypothetical room. The left plot shows the execution of the EM algorithm and the classification of each positioning estimator.

As it is possible to notice, Equation (6.58) is intractable because of the computational complexity of the optimization procedure. Thus, it is common to approach the problem in an iterative fashion in which we assign elements to groups and then determine the mean value of the elements in the group. Given a set of initial mean values for each group $\{m_1^{(0)}, \ldots, m_k^{(0)}\}$ (the superscript index identifies the iteration step) defined in a random or heuristic manner, the iterative k-means algorithm can be defined as follows.

Assigning: The Euclidean distance is calculated between all group mean values and each element, so that each element is assigned to the group to which it has the shortest Euclidean distance. This decomposition is known as the Voronoi diagram:

$$X_j \in \hat{G}_i^{(l)} = \arg\min_{G_i^{(l)}} \|X_j, m_i^{(l)}\|, \tag{6.59}$$

where l is the iteration step.

Updating: The mean values for each group at time step $l+1$ are updated to $\{m_1^{(l+1)}, \ldots, m_k^{(l+1)}\}$ according to the average of all the elements in the group:

$$m_i^{(l+1)} = \frac{1}{n_i^{(l)}} \sum_{X_j \in G_i^{(l)}} X_j, \tag{6.60}$$

where $n_i^{(l)}$ is the number of elements in group $G_i^{(l)}$ at time step l.

Once the assignment of elements no longer changes in subsequent time steps, the algorithm terminates. As in any other optimization algorithm, convergence to an absolute minimum is not assured.

The k-means algorithm can be used, for instance, to estimate average routes obtained from several tracking estimations performed in a predetermined area. Let

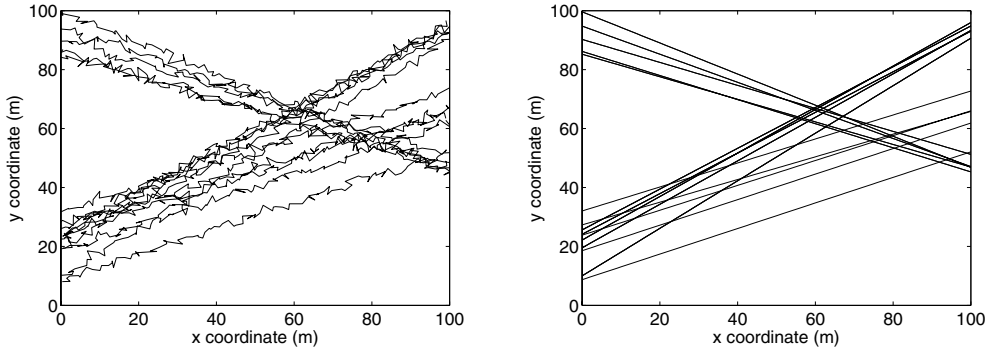


Figure 6.16 Example of 15 noisy routes (left) and the corresponding linear regression (right).

us assume, for example, a scenario with three different routes. For each route, there are five estimated paths, supposedly obtained from a tracking algorithm. The actual tracking algorithm is unimportant in this example, since our interest is focused on the k-means algorithm for grouping estimated trajectories. Let us suppose that these 15 different estimated paths are those given in Figure 6.16 (left). As one can see, the estimated paths may be corrupted by noise. Each of the 15 estimated trajectories is composed of a sequence of position coordinates, which typically differ in number from another estimated trajectory. This problem makes it complex to directly establish a definition of a distance between two trajectories. Hence, we have determined for each trajectory the corresponding linear regression, which is plotted in Figure 6.16 (right). This regression permits the approximation of the trajectory by a known expression so that it is possible to calculate distances between different estimated trajectories. In order to calculate the distances between pairs of trajectories, all the expressions obtained from the regression were sampled at equidistant points, which were used to calculate the mean square error. This metric was defined as the distance between trajectories and was used in the assignment step to calculate the distance to the center of the group, which itself had an initially random trajectory. By applying the k-means algorithm given above, it was possible to determine three routes as given in Figure 6.17. As Figure 6.17 shows, the k-means algorithm was able to approximately determine the three routes that existed.

6.5 Conclusions

The chapter started by explaining the increase in complexity introduced into positioning systems when movement of MSs and then cooperation among MSs are additionally considered. The chapter then described some of the most common mobility models used as a fundamental part of wireless positioning, in particular for simulation purposes. These models concern individual mobility, mobility on maps,

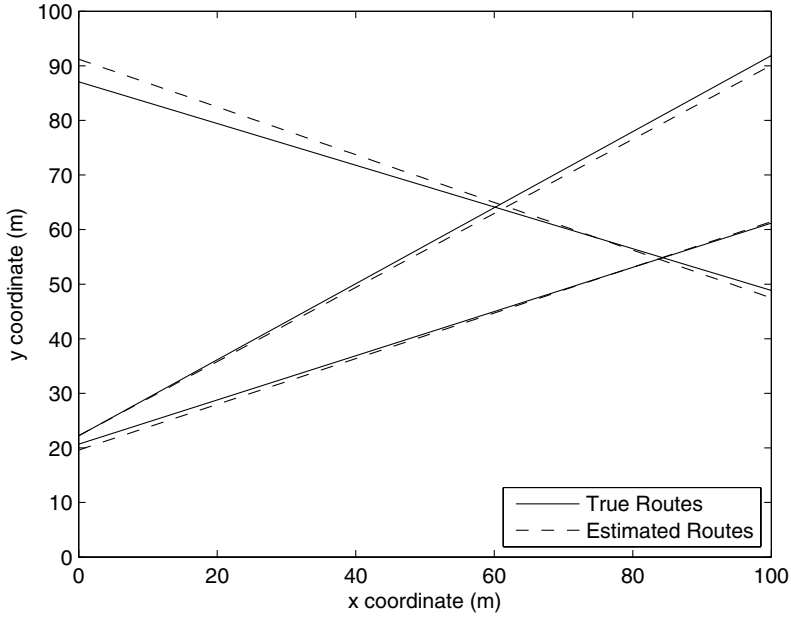


Figure 6.17 True routes and routes estimated using the k-means algorithm.

group mobility and social-based mobility models. The next section presented the various steps involved in a tracking system. In an initial step, the measurements are processed in order to filter out some useless measurements. Then, algorithms for nonmaneuvering and maneuvering MSs were applied. Finally, envisioning cooperative schemes in wireless positioning, we described some methods and techniques for learning position and path patterns.

7

Error Mitigation Techniques

Ismail Guvenc

Wireless Access Laboratory, DOCOMO Communications Laboratories, Palo Alto, CA, USA

7.1 Introduction

The previous chapters have introduced several algorithms and mechanisms to perform positioning and tracking of MSs in wireless networks. The common characteristic of all those methods is that all the approaches considered use raw measurements of the wireless channel, i.e., no preprocessing of the measurements is done to detect possible destructive effects existing in the measured data. Thus, this chapter describes the mechanisms that can mitigate and circumvent some of these destructive effects present in raw measurements.

The most common effect affecting position accuracy is the NLOS effect. NLOS scenarios occur when there is an obstruction between the transmitter (TX) and the receiver (RX) and are commonly encountered in modern wireless system deployment in both indoor environments (e.g., residential and office buildings shopping malls) and outdoor environments (e.g., metropolitan and urban areas). As illustrated in Figure 7.1, the signal transmitted by the TX arrives at the RX through multiple paths. In an NLOS scenario, the LOS path is blocked by an obstruction and hence does not reach the RX.[1] Therefore, even if the first path arriving at the RX can be correctly identified, the distance estimate calculated using this path includes a

[1] In some scenarios, a signal following this path may also arrive at the receiver after being attenuated and delayed by the obstruction (Allen et al. 2007).

Mobile Positioning and Tracking João Figueiras and Simone Frattasi

© 2010 John Wiley & Sons, Ltd

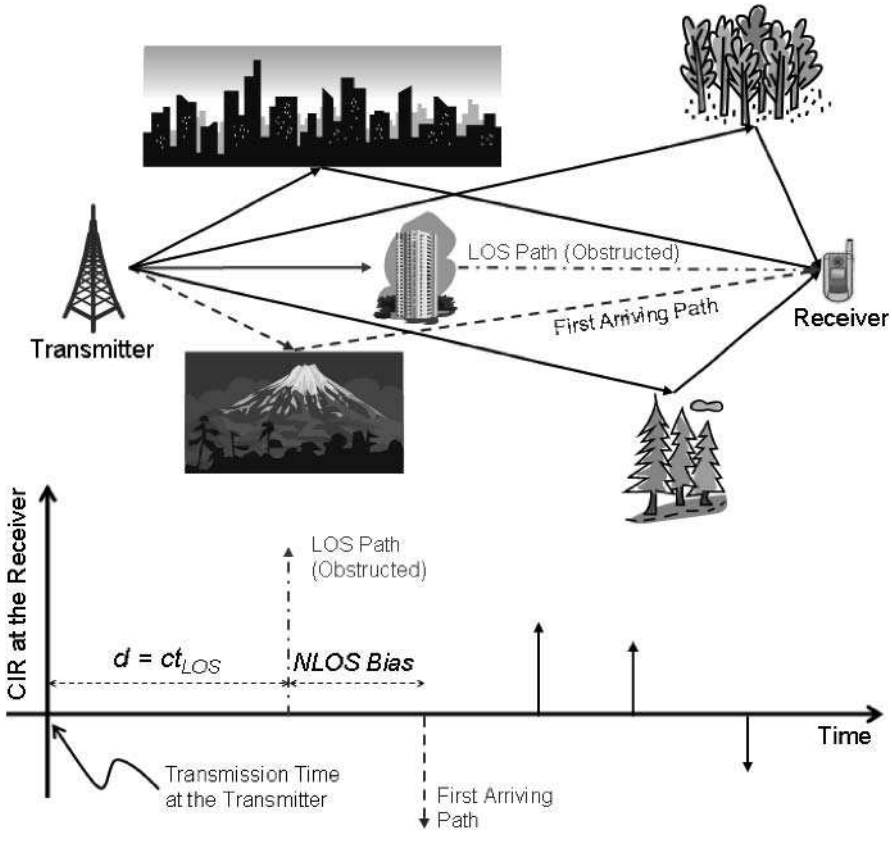


Figure 7.1 Illustration of the NLOS problem in wireless localization.

positive bias, which, if not handled properly in the localization algorithm, may considerably degrade the localization accuracy.

Several previous studies reported in the literature (e.g., Gezici 2008; Gustafsson and Gunnarsson 2005a; Liu et al. 2007; Patwari et al. 2005; Sayed et al. 2005b) provide extensive reviews of localization techniques using AOA, TOA, TDOA and RSS. However, the impact of NLOS propagation and how this can be mitigated has not been investigated in detail. In Caffery (1999), an analysis of NLOS effects in CDMA cellular networks was provided and two particular NLOS mitigation techniques were summarized: LOS reconstruction and NLOS measurement weighting. However, several other NLOS mitigation techniques have been proposed in the literature subsequently. The goal of the current chapter is to provide a comprehensive and up-to-date survey of NLOS mitigation techniques for wireless location estimation. While this chapter will focus on TOA-based approaches for the sake of a unified analysis, several of the NLOS mitigation techniques

reviewed may also be applied to TDOA-, AOA- and RSS-based systems after some simple extensions.

The chapter is organized as follows. Section 7.2 extends previous chapters by outlining the TOA-based localization scenario and presenting the system model and the problem definition to be used throughout the chapter. A brief overview of the fundamental lower bounds in LOS scenarios is provided, and some of the key techniques based on the maximum-likelihood (ML) method, which can achieve accuracies close to these fundamental bounds, are summarized. Section 7.3 provides lower bounds and some ML techniques for localization in NLOS scenarios. Section 7.4 is on LS techniques, Section 7.5 on constraint-based techniques, Section 7.6 is on robust estimators and Section 7.7 is a review of the identify and discard type of techniques, all designed for improving the localization accuracy in NLOS scenarios. Finally, Section 7.8 briefly summarizes the techniques discussed and provides some concluding remarks.

7.2 System Model

Consider a wireless network as shown in Figure 7.2 where there are n FRPs,[2] $\hat{X} = [\hat{x}\ \ \hat{y}]^{\mathrm{T}}$ is the estimate of the MS location, and $\hat{d}_i$ is the measured distance between the MS and the ith FRP:

$$\hat{d}_i = d_i + b_i + \eta_i = ct_i, \quad i = 1, 2, \ldots, n, \tag{7.1}$$

where t_i is the TOA of the signal at the ith FRP, c is the speed of light, d_i is the real distance between the MS and the ith FRP, $\eta_i \sim \mathrm{Norm}(0, \sigma_i^2)$ is the additive white Gaussian noise (AWGN), with variance σ_i^2, and b_i is a positive distance bias introduced owing to the blockage of direct path, given by

$$b_i = \begin{cases} 0 & \text{if the } i\text{th FRP is in LOS,} \\ \psi_i & \text{if the } i\text{th FRP is in NLOS.} \end{cases} \tag{7.2}$$

For NLOS FRPs, the bias term ψ_i has been modeled in different ways in the literature depending on the wireless propagation channel environment; it may be exponentially distributed (Chen 1999; Gezici and Sahinoglu 2004), uniformly distributed (Jourdan and Roy 2006; Venkatesh and Buehrer 2006), Gaussian distributed (Dizdarevic and Witrisal 2006); constant in a time window (Riba and Urruela 2004) or based on an empirical model derived from measurements (Denis et al. 2006; Jourdan et al. 2006).

Let

$$D = D(X) = [d_1, d_2, \ldots, d_n]^{\mathrm{T}} \tag{7.3}$$

be a vector of true distances between the MS and the n FRPs, let

$$\hat{D} = [\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_n]^{\mathrm{T}} \tag{7.4}$$

[2]The FRPs are usually BSs in cellular networks, anchor nodes in sensor networks, and APs in WLANs.

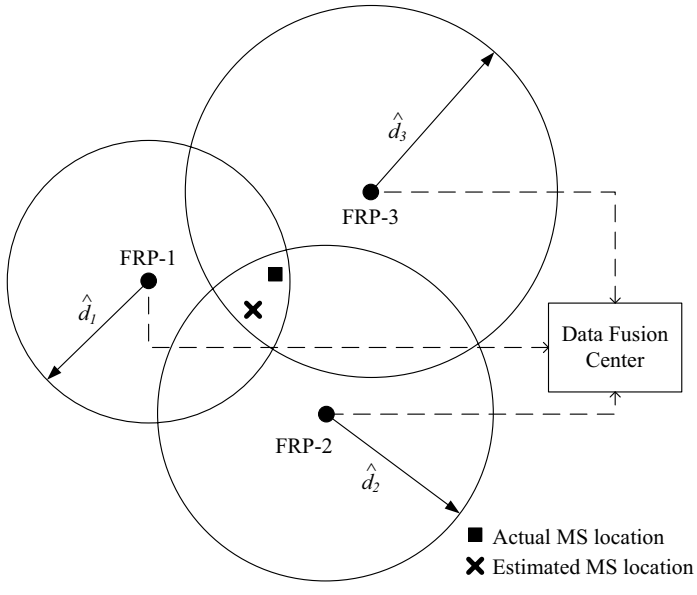


Figure 7.2 Illustration of a simple scenario for wireless localization.

be a vector of measured distances and let

$$B = [b_1, b_2, \ldots, b_n]^{\mathrm{T}} \tag{7.5}$$

be a bias vector. Also, let

$$Q = \mathrm{E}[\eta\eta^{\mathrm{T}}] = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2]^{\mathrm{T}} \tag{7.6}$$

denote the covariance of the noise vector $\eta = [\eta_1, \eta_2, \ldots, \eta_n]^{\mathrm{T}}$ with the assumption that all of the noise terms are zero-mean, independent Gaussian random variables.

In the absence of noise and NLOS bias, the true distance d_i between the MS and the ith FRP defines a circle around the ith FRP corresponding to possible MS locations:

$$(x - x_i)^2 + (y - y_i)^2 = d_i^2, \quad i = 1, 2, \ldots, n \tag{7.7}$$

where all the circles intersect at the same point, and solving these equations jointly gives the exact MS location. However, the noisy measurements and NLOS bias at different FRPs yield circles which do not intersect at the same point, resulting in inconsistent equations as follows:

$$(x - x_i)^2 + (y - y_i)^2 = \hat{d}_i^2, \quad i = 1, 2, \ldots, n. \tag{7.8}$$

In order to have more compact expressions throughout the chapter, we also define the following terms:

$$s = x^2 + y^2 \quad \text{and} \quad k_i = x_i^2 + y_i^2. \tag{7.9}$$

The problem of TOA-based location estimation can be defined as the estimation of the location X of an MS from the noisy (and possibly biased) distance measurements in Equation (7.4) given the FRP locations X_i; in other words, given the set of equations in Equation (7.8). Various localization techniques have been proposed in the literature to estimate the MS location from Equation (7.8) in LOS scenarios (which have also been discussed in Chapter 5) and NLOS scenarios. Before reviewing the various NLOS mitigation algorithms, first we will briefly review ML-based localization techniques and fundamental lower bounds in LOS scenarios.

7.2.1 Maximum-likelihood algorithm for LOS scenarios

In the absence of NLOS bias (i.e., $b_i = 0$ for all i), the conditional PDF of $\hat{D}$ in Equation (7.4) given X can be expressed as follows (Chan et al. 2006a; Cheng and Sahai 2004):

$$\mathrm{f}(\hat{D}|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(\hat{d}_i - d_i)^2}{2\sigma_i^2}\right\} \tag{7.10}$$

$$= \frac{1}{\sqrt{(2\pi)^n \det(Q)}} \exp\left\{-\frac{J}{2}\right\}, \tag{7.11}$$

where

$$J = [\hat{D} - D(X)]^{\mathrm{T}} Q^{-1} [\hat{D} - D(X)], \tag{7.12}$$

with Q as given in Equation (7.6). Then, the ML solution for X is the one that maximizes $\mathrm{f}(\hat{D}|X)$, i.e.,

$$\hat{X} = \arg\max_{X} \mathrm{f}(\hat{D}|X). \tag{7.13}$$

Note that solving for X from Equation (7.13) requires a search over possible MS locations, which is computationally intensive.

For the special case of $\sigma_i^2 = \sigma^2$ for all i, the ML solution of Equation (7.13) is equivalent to minimizing J. In order to find the minimum value of J, the gradient of J with respect to X is equated to zero, yielding (Chan et al. 2006a)

$$\sum_{i=1}^{n} \frac{(d_i - \hat{d}_i)(x - x_i)}{d_i} = 0 \tag{7.14}$$

and

$$\sum_{i=1}^{n} \frac{(d_i - \hat{d}_i)(y - y_i)}{d_i} = 0, \tag{7.15}$$

which are nonlinear equations. Hence, X cannot be solved for in closed form from Equations (7.14) and (7.15) using a linear LS algorithm. Also, both Equation (7.14) and Equation (7.15) depend on the d_i, which are unknown. However, even though a closed-form ML solution is not possible, iterative ML and approximate ML techniques can be derived as discussed in Chan et al. (2006a) and Chan and Ho (1994), respectively, which may asymptotically achieve the Cramér–Rao lower bound (CRLB).

7.2.2 Cramér–Rao lower bounds for LOS scenarios

Given the conditional PDF of $\hat{D}$ as in Equation (7.10), we may derive the CRLB for TOA-based location estimation (see Section 4.6.3). The CRLB, in general, can be defined as the theoretical lower bound on the variance of any unbiased estimator of an unknown parameter. The CRLB for TOA-based location estimation depends mainly on the following factors: (1) the positions of the FRPs (X_i); (2) the true position of the MS (X); and (3) the measurement noise variances (σ_i^2). It is calculated using the Fisher information matrix (FIM), whose elements are defined as

$$[I(X)]_{ij} = -\mathrm{E}\left[\frac{\partial^2 \ln \mathrm{f}(\hat{D}|X)}{\partial X_i \, \partial X_j}\right]. \tag{7.16}$$

Then, using the PDF given in Equation (7.10), the FIM can be calculated as (Chan et al. 2006b; Cheng and Sahai 2004)

$$I(X) = \begin{bmatrix} \sum_{i=1}^{n} \frac{(x-x_i)^2}{\sigma_i^2 d_i^2} & \sum_{i=1}^{n} \frac{(x-x_i)(y-y_i)}{\sigma_i^2 d_i^2} \\ \sum_{i=1}^{n} \frac{(x-x_i)(y-y_i)}{\sigma_i^2 d_i^2} & \sum_{i=1}^{n} \frac{(y-y_i)^2}{\sigma_i^2 d_i^2} \end{bmatrix} \tag{7.17}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \frac{\cos^2(\theta_i)}{\sigma_i^2} & \sum_{i=1}^{n} \frac{\cos(\theta_i)\sin(\theta_i)}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{\cos(\theta_i)\sin(\theta_i)}{\sigma_i^2} & \sum_{i=1}^{n} \frac{\sin^2(\theta_i)}{\sigma_i^2} \end{bmatrix}, \tag{7.18}$$

where θ_i defines the angle from the ith FRP to the MS, and the CRLB is given by $I^{-1}(X)$. Thus, for an estimate $\hat{X}$ of the MS location obtained with any unbiased estimator, we have

$$\mathrm{E}_X[(\hat{X} - X)(\hat{X} - X)^{\mathrm{T}}] \geq I^{-1}(X). \tag{7.19}$$

The CRLB can be related to another important measurement metric, referred to as the geometric dilution of precision (GDOP) in Section 4.6.2. For identical noise variances $\sigma_i^2 = \sigma^2$ at different FRPs, the GDOP can be defined as

$$\mathrm{GDOP} = \frac{\mathrm{RMSE}_{\mathrm{loc}}}{\mathrm{RMSE}_{\mathrm{range}}} = \frac{\sigma_{\mathrm{loc}}}{\sigma}, \tag{7.20}$$

where $\mathrm{RMSE}_{\mathrm{loc}}$ and $\mathrm{RMSE}_{\mathrm{range}}$ are the root mean square error (RMSE) of the location estimate and the range estimate, respectively, and σ_{loc} is the standard deviation of the location estimate. Similarly to the CRLB, the GDOP also depends on the positions of the FRP and the MS; however, it is independent of the measurement noise variance σ_i^2. GDOP values smaller than three are usually preferable, and values larger than six may imply a very bad geometry of the FRPs. If the location estimator employed can

achieve the CRLB, the GDOP is given by

$$\text{GDOP} = \frac{\sqrt{\text{trace}[I^{-1}(X)]}}{\sigma} \tag{7.21}$$

$$= \sqrt{\text{trace}[\tilde{I}^{-1}(X)]}, \tag{7.22}$$

where

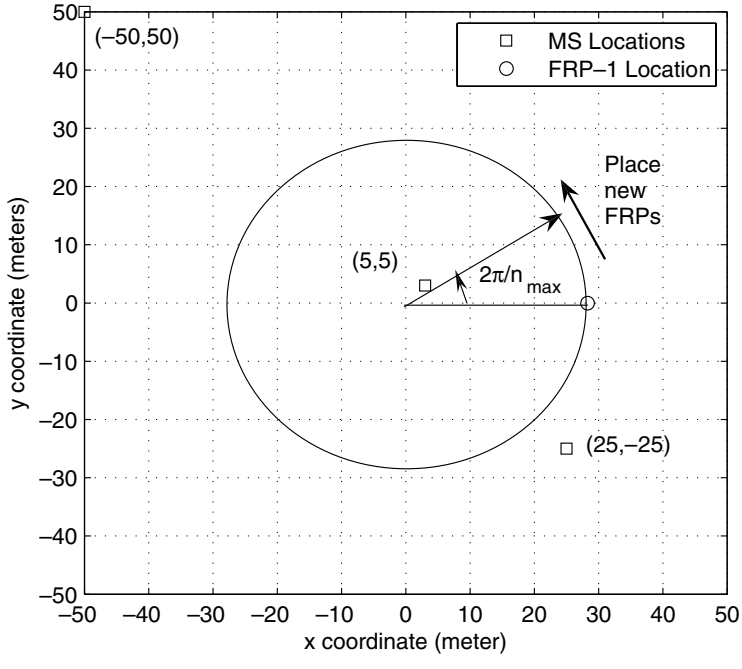
$$\tilde{I}(X) = \begin{bmatrix} \sum_{i=1}^{n} \cos^2(\theta_i) & \sum_{i=1}^{n} \cos(\theta_i)\sin(\theta_i) \\ \sum_{i=1}^{n} \cos(\theta_i)\sin(\theta_i) & \sum_{i=1}^{n} \sin^2(\theta_i) \end{bmatrix}. \tag{7.23}$$

The relation between the achievable localization accuracy and the geometry between the locations of the MS and the FRPs is evident from Equation (7.22).

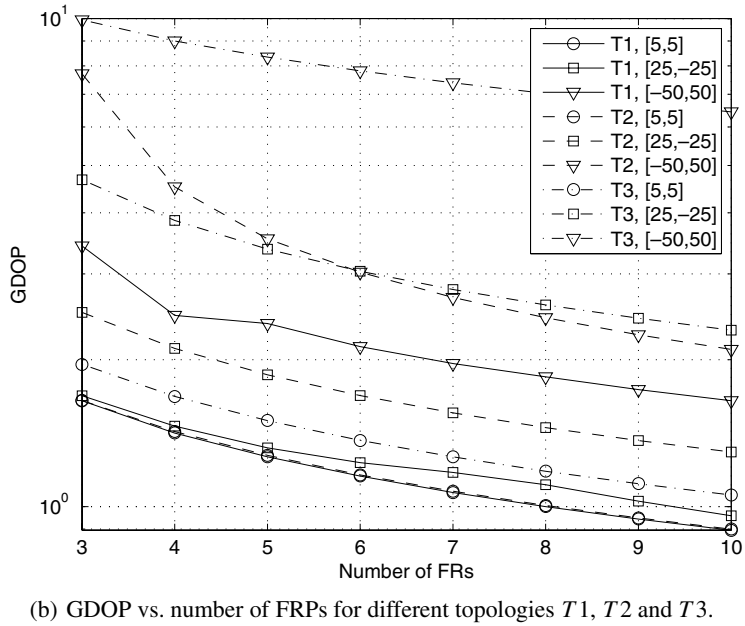
Example 7.1 *In order to see some typical values of the GDOP in wireless location estimation, some simple node topologies were simulated as shown in Figure 7.3. Three MS locations were considered, namely* $[5, 5]$ *m,* $[25, -25]$ *m and* $[-50, 50]$ *m. The location of FRP-1 was fixed at* $[20\sqrt{2}, 0]$ *m and new FRPs were added counterclockwise around the circle in the figure. Three topologies were considered, namely T1, T2 and T3, in which the increment was* $2\pi/n_{\max}$, $\pi/n_{\max}$ *and* $\pi/2n_{\max}$ *radians, respectively, where* $n_{\max} = 10$ *denotes the maximum number of FRPs. The GDOP was less than two for an MS located at* $[5, 5]$ *m for all of these topologies, and became better the more FRPs were deployed. For an MS located at* $[-50, 50]$ *m, the GDOP was the worst, and it could be as large as* 10 *for T3. The results show that, as we increase the number of FRPs, it is possible to have GDOP values smaller than* 1*, which implies that the standard deviation of the location estimate may become smaller than the standard deviation of the distance estimates for well-designed topologies.*

7.3 NLOS Scenarios: Fundamental Limits and ML Solutions

In realistic situations, it may be possible that the LOS between the MS and some of the FRPs may be obstructed (i.e., $b_i > 0$ for some i). These NLOS FRPs may seriously degrade the localization accuracy. The simplest way to mitigate the effects of NLOS is achieved by identifying and discarding the NLOS FRPs and estimating the MS location by using one of the LOS techniques discussed in the previous section. However, there is always the possibility of false alarms (identifying an LOS FRP as NLOS) and missed detections (identifying an NLOS FRP as LOS), which degrade the localization accuracy. In this section, we review alternative NLOS mitigation techniques reported in the literature. First, the ML-based techniques and the CRLBs in NLOS scenarios will be discussed.



(a) Topology for GDOP simulations.



(b) GDOP vs. number of FRPs for different topologies $T1$, $T2$ and $T3$.

Figure 7.3 GDOPs for different topologies and MS locations (Guvenc and Chong 2009).

7.3.1 ML-based algorithms

ML approaches to NLOS mitigation were discussed in Gezici and Sahinoglu (2004) and Riba and Urruela (2004); these approaches require prior knowledge regarding the distribution of NLOS bias. For example, Gezici and Sahinoglu (2004) provided an ML solution for the position of the MS with the assumption that the NLOS bias b_i is exponentially distributed with parameter λ_i. Since in most cases the NLOS bias is much larger than the Gaussian measurement error, the following simplifying assumption was made:

$$\hat{d}_i = d_i + \begin{cases} b_i, & i = 1, 2, \ldots, n_{\mathrm{NL}}, \\ \eta_i, & i = n_{\mathrm{NL}} + 1, \ldots, n, \end{cases} \tag{7.24}$$

where n_{NL} is the number of NLOS FRPs and $n_{\mathrm{L}} = n - n_{\mathrm{NL}}$ is the number of LOS FRPs. Thus, a simplified ML solution is given as follows (Gezici and Sahinoglu 2004):

$$\hat{X} = \arg\min_{X} \left\{ \sum_{i=1}^{n_{\mathrm{NL}}} \lambda_i (\hat{d}_i - d_i) + \sum_{i=n_{\mathrm{NL}}+1}^{n} \frac{(\hat{d}_i - d_i)^2}{2\sigma_i^2} \right\}. \tag{7.25}$$

It is also possible to obtain an exact decision rule by considering a summation of Gaussian and exponential random variables, which has the following probability density function:

$$\mathrm{f}(u) = \lambda \exp\left(-\lambda \left(u - \frac{\lambda \sigma^2}{2} \right) \right) Q\left(\lambda\sigma - \frac{u}{\sigma} \right), \tag{7.26}$$

where $Q(a) = (1/\sqrt{2\pi}) \int_a^{\infty} \exp(-u^2/2)\, \mathrm{d}u$ denotes the Q-function, and the exact ML solution becomes (Gezici and Sahinoglu 2004)

$$\hat{X} = \arg\min_{X} \left\{ \sum_{i=1}^{n_{\mathrm{NL}}} \lambda_i \left(\hat{d}_i - d_i - \frac{\lambda_i \sigma_i^2}{2} \right) - \sum_{i=1}^{n_{\mathrm{NL}}} \log\left[Q\left(\lambda_i \sigma_i - \frac{\hat{d}_i - d_i}{\sigma_i} \right) \right] + \sum_{i=n_{\mathrm{NL}}+1}^{n} \frac{(\hat{d}_i - d_i)^2}{2\sigma_i^2} \right\}. \tag{7.27}$$

Example 7.2 *Consider a simple cellular wireless localization scenario with seven FRPs as illustrated in Figure 7.4. An MS is located at* (1, 1.5) *km and the locations of the FRPs are as illustrated in the figure. Each of the cells has a radius of* 3 *km and it is assumed that the transmitted MS signal is successfully received at all seven FRPs and utilized for location estimation. The distance estimate at each of the FRPs is subject to zero-mean AWGN with a variance of* 0.2 km^2. *Moreover, for distance estimates at FRP-1 an exponentially distributed NLOS bias exists, which is assumed to have a mean equivalent to* 10% *of the true measurement. Given this scenario, the metrics for the approximate ML in Equation* (7.25) *and exact ML in Equation* (7.27) *are plotted in Figures 7.4(a) and 7.4(b), respectively, for one realization of the noise and NLOS bias. It is observed that both metrics are minimized close to the true location of the MS.*

Another NLOS mitigation technique for TOA-based systems based on the ML approach was introduced by Riba and Urruela (2004). In this technique, several hypotheses about the set of FRPs are considered, and then, utilizing the ML principle, the best set (which is assumed to be composed of LOS FRPs) is selected for location estimation. The hypothesis index estimate for the best FRP set is derived as (Riba and Urruela 2004)

$$\hat{i} = \arg\min_{i} \left[\ln \gamma^{-1}(i) + \sum_{j \in S_i^{\mathrm{LOS}}} \frac{n_{\mathrm{trn}}}{2} \frac{\hat{\sigma}_j^2(i)}{\sigma_j^2} \right], \tag{7.28}$$

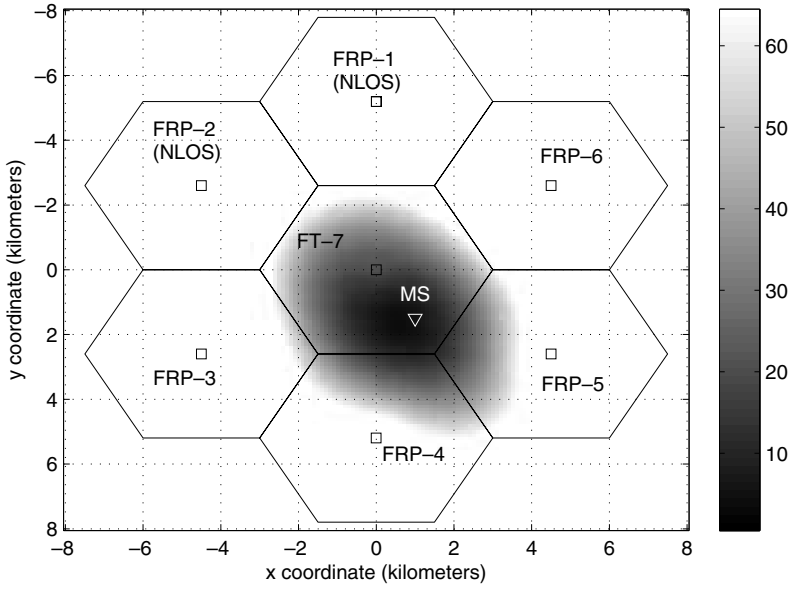
where S_i^{LOS} denotes the ith set of FRPs which are hypothesized to be in LOS, the $\gamma(i)$ are assigned according to the a priori probability of each hypothesis (equivalent to 1, if no information is available),

$$\hat{\sigma}_j^2(i) = \frac{1}{n_{\mathrm{trn}}} \sum_{k=1}^{n_{\mathrm{trn}}} \left(t_j^{(k)} - \frac{\| \hat{X}(i) - X_k \|}{c} \right)^2 \tag{7.29}$$

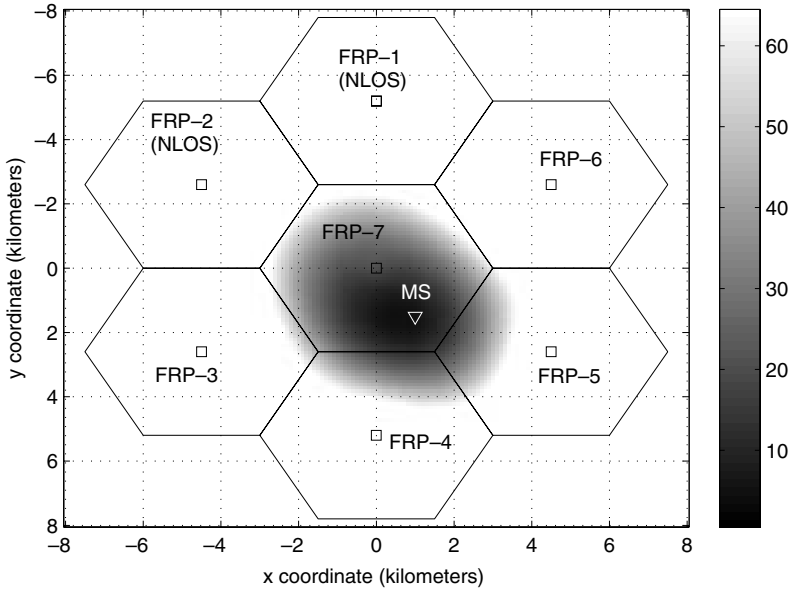
denotes the estimated variance of the n_{trn} TOA measurements associated with the jth FRP under the ith hypothesis, and $t_j^{(k)}$ denotes the kth TOA measurement at the jth FRP. Note that the above approach requires buffering of n_{trn} TOA measurements for the purpose of obtaining the noise statistics for a particular FRP. Once the set of LOS FRPs has been selected using the ML principle, the MS location is estimated using only these FRPs and the ML algorithm. Simulation results (Riba and Urruela 2004) show that this yields better accuracy than the residual-weighting (Rwgh) technique introduced by Chen (1999) (see Section 7.4.2), and slightly worse than when only the true LOS FRPs are used for localization. Also, at low SNR and small NLOS bias values, simulation results show that it may be better not to employ any NLOS mitigation, in order not to degrade the accuracy.

7.3.2 Cramér–Rao lower bound

In NLOS scenarios, the CRLB depends on whether there is any prior information available about the NLOS bias. First, let us assume that there is no prior information



(a) Approximate ML metric in Equation (7.25).



(b) Exact ML metric in Equation (7.27).

Figure 7.4 Simulations of exact and approximate ML techniques for NLOS mitigation.

about the NLOS bias, but that we know the FRPs that are in NLOS. Then, an extended version of the FIM in Equation (7.19) is given by (Qi et al. 2006)

$$I(X_b) = \tilde{A} I(D) \tilde{A}^{\mathrm{T}}, \tag{7.30}$$

where

$$X_b = [x,\ y,\ b_1,\ b_2, \ldots, b_{n_{\mathrm{NL}}}]^{\mathrm{T}} \tag{7.31}$$

is an $(n_{\mathrm{NL}} + 2) \times 1$ vector of unknown parameters incorporating the NLOS bias values, and

$$\tilde{A} = \frac{\partial D}{\partial X_b} = \begin{bmatrix} \frac{\partial d_1}{\partial x} & \frac{\partial d_2}{\partial x} & \cdots & \frac{\partial d_{n_{\mathrm{NL}}}}{\partial x} & \cdots & \frac{\partial d_n}{\partial x} \\ \frac{\partial d_1}{\partial y} & \frac{\partial d_2}{\partial y} & \cdots & \frac{\partial d_{n_{\mathrm{NL}}}}{\partial y} & \cdots & \frac{\partial d_n}{\partial y} \\ \frac{\partial d_1}{\partial b_1} & \frac{\partial d_2}{\partial b_1} & \cdots & \frac{\partial d_{n_{\mathrm{NL}}}}{\partial b_1} & \cdots & \frac{\partial d_n}{\partial b_1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial d_1}{\partial b_{n_{\mathrm{NL}}}} & \frac{\partial d_2}{\partial b_{n_{\mathrm{NL}}}} & \cdots & \frac{\partial d_{n_{\mathrm{NL}}}}{\partial b_{n_{\mathrm{NL}}}} & \cdots & \frac{\partial d_n}{\partial b_{n_{\mathrm{NL}}}} \end{bmatrix}, \tag{7.32}$$

and

$$I(D) = \mathrm{E}_D \left[\frac{\partial \ln f(\hat{D}|D)}{\partial D} \left(\frac{\partial \ln f(\hat{D}|D)}{\partial D} \right)^{\mathrm{T}} \right]. \tag{7.33}$$

As discussed in Qi et al. (2006), $\tilde{A}$ can be written in terms of its LOS and NLOS components as

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{\mathrm{NL}} & \tilde{A}_{\mathrm{L}} \\ I_{n_{\mathrm{NL}}} & \mathbf{0}_{n_{\mathrm{NL}}, n_{\mathrm{L}}} \end{bmatrix}, \tag{7.34}$$

where $I_{n_{\mathrm{NL}}}$ and $\mathbf{0}_{n_{\mathrm{NL}}, n_{\mathrm{L}}}$ are an identity matrix of size $n_{\mathrm{NL}} \times n_{\mathrm{NL}}$ and a zero matrix of size $n_{\mathrm{NL}} \times n_L$, respectively, and

$$\tilde{A}_{\mathrm{NL}} = \begin{bmatrix} \cos\theta_1 & \cos\theta_2 & \ldots & \cos\theta_{n_{\mathrm{NL}}} \\ \sin\theta_1 & \sin\theta_2 & \ldots & \sin\theta_{n_{\mathrm{NL}}} \end{bmatrix} \tag{7.35}$$

and

$$\tilde{A}_{\mathrm{L}} = \begin{bmatrix} \cos\theta_{n_{\mathrm{NL}}+1} & \cos\theta_{n_{\mathrm{NL}}+2} & \ldots & \cos\theta_n \\ \sin\theta_{n_{\mathrm{NL}}+1} & \sin\theta_{n_{\mathrm{NL}}+2} & \ldots & \sin\theta_n \end{bmatrix}. \tag{7.36}$$

Similarly, $I(D)$ can be written in terms of its NLOS and LOS components as (Qi et al. 2006)

$$I(D) = \begin{bmatrix} \Lambda_{\mathrm{NL}} & \mathbf{0} \\ \mathbf{0} & \Lambda_{\mathrm{L}} \end{bmatrix}, \tag{7.37}$$

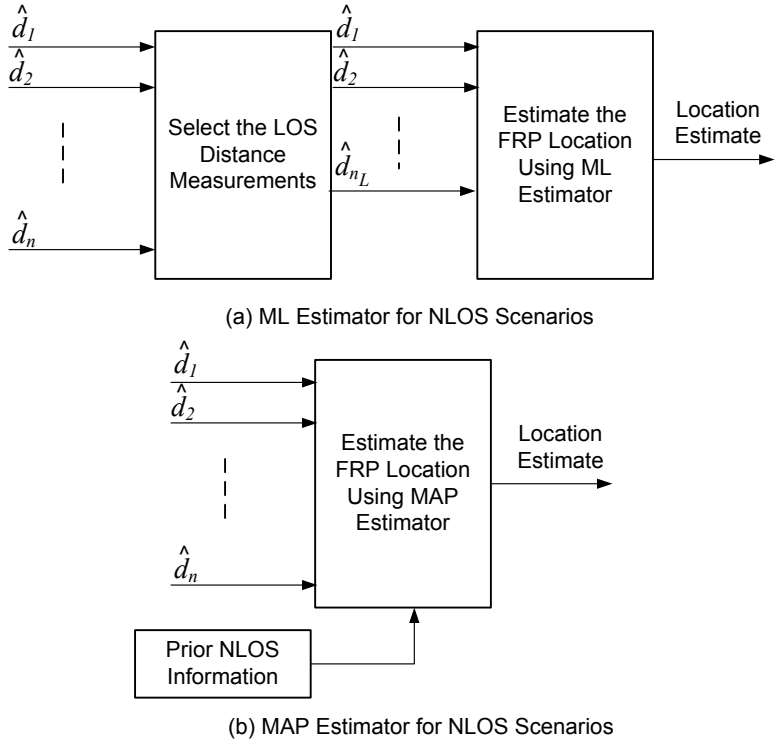


Figure 7.5 Block diagrams for (a) ML estimator and (b) MAP estimator for NLOS scenarios. In (a), without loss of generality, it is assumed that the first n_{L} measurements are the LOS measurements (Guvenc and Chong 2009).

where $\Lambda_{\mathrm{NL}} = \mathrm{diag}(\sigma_1^{-2}, \ldots, \sigma_{n_{\mathrm{NL}}}^{-2})$ and $\Lambda_{\mathrm{L}} = \mathrm{diag}(\sigma_{n_{\mathrm{NL}}+1}^{-2}, \ldots, \sigma_n^{-2})$. After some manipulations, $I(X_b)$ can be obtained as (Qi et al. 2006)

$$I(X_b) = \begin{bmatrix} \tilde{A}_{\mathrm{NL}} \Lambda_{\mathrm{NL}} \tilde{A}_{\mathrm{NL}}^{\mathrm{T}} + \tilde{A}_{\mathrm{L}} \Lambda_{\mathrm{L}} \tilde{A}_{\mathrm{L}}^{\mathrm{T}} & \tilde{A}_{\mathrm{NL}} \Lambda_{\mathrm{NL}} \\ \Lambda_{\mathrm{NL}} \tilde{A}_{\mathrm{NL}}^{\mathrm{T}} & \Lambda_{\mathrm{NL}} \end{bmatrix}. \tag{7.38}$$

Note that Equation (7.38) depends on both NLOS and LOS signals. However, it was proven in Qi et al. (2006) that the CRLB for the MS location is given by

$$\mathrm{E}[(\hat{X}_b - X_b)(\hat{X}_b - X_b)^{\mathrm{T}}] \geq I^{-1}(X_b) = (\tilde{A}_{\mathrm{L}} \Lambda_{\mathrm{L}} \tilde{A}_{\mathrm{L}}^{\mathrm{T}})^{-1}. \tag{7.39}$$

In other words, the CRLB depends exclusively on the LOS signals if the NLOS FRPs can be accurately identified. Hence, an ML estimator that can achieve the CRLB in NLOS scenarios first identifies the NLOS FRPs and discards their measurements, and then obtains the location estimate using the LOS FRPs (Figure 7.5(a)).

If there is further side information related to the statistics of the NLOS bias vector B, a better positioning accuracy can be obtained. Then, the generalized Cramer–Rao

lower bound (G-CRLB) can be written as (Qi et al. 2006)

$$\mathrm{E}[(\hat{X}_b - X_b)(\hat{X}_b - X_b)^{\mathrm{T}}] \geq \left(I(X_b) + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \right)^{-1} \tag{7.40}$$

$$= \left(\begin{bmatrix} \tilde{A}_{\mathrm{NL}} \Lambda_{\mathrm{NL}} \tilde{A}_{\mathrm{NL}}^{\mathrm{T}} + \tilde{A}_{\mathrm{L}} \Lambda_{\mathrm{L}} \tilde{A}_{\mathrm{L}}^{\mathrm{T}} & \tilde{A}_{\mathrm{NL}} \Lambda_{\mathrm{NL}} \\ \Lambda_{\mathrm{NL}} \tilde{A}_{\mathrm{NL}}^{\mathrm{T}} & \Lambda_{\mathrm{NL}} + \boldsymbol{\Omega} \end{bmatrix} \right)^{-1}, \tag{7.41}$$

where $\boldsymbol{\Omega} = \mathrm{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_{n_{\mathrm{NL}}}^2)$, and $\tilde{\sigma}_i^2$ can be interpreted as the variance[3] of b_i. As an upper bound on the G-CRLB, when the variances $\tilde{\sigma}_i^2$ are infinitely large, the G-CRLB is reduced to the CRLB since there is practically no information available about b_i. The estimator that asymptotically achieves the G-CRLB is given by the maximum a posteriori (MAP) estimator, and it employs the statistics of the NLOS biases (Figure 7.5(b)).

7.4 Least-squares Techniques for NLOS Localization

The LS techniques for location estimation can be tuned to suppress the effects of NLOS bias, for example through some appropriate weighting. In this section, weighted least-squares approaches and the residual-weighting algorithm will be briefly reviewed.

7.4.1 Weighted least squares

A simple way to mitigate the effects of NLOS FRPs is to give less emphasis to the corresponding NLOS terms in the LS solution. In Caffery and Stüber (1998) and Gezici and Sahinoglu (2004), with the assumption that the variances of the distance measurements are larger for NLOS FRPs, the inverses of these variances are used as a reliability metric. From the ML algorithm, the location estimate is given by

$$X_{\mathrm{ML}} = \arg \max_X \mathrm{p}(\hat{D}|X), \tag{7.42}$$

where

$$\mathrm{f}(\hat{D}|X) = \mathrm{f_n}(\hat{D} - D|X). \tag{7.43}$$

If the noise is Gaussian distributed, we have

$$\mathrm{f_n}(\eta) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\eta^2}{2\sigma_i^2} \right). \tag{7.44}$$

Then, the joint probability function becomes

$$\mathrm{f}(\hat{D}|X) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} \sigma_i} \exp\left(-\sum_{i=1}^{n} \frac{(\hat{d}_i - \|X - X_i\|)^2}{2\sigma_i^2} \right). \tag{7.45}$$

[3] For a Gaussian-distributed NLOS bias, this is strictly the variance of the NLOS bias.

Upon further manipulation of Equation (7.45), the ML solution becomes equivalent to

$$\hat{X}_{\mathrm{ML}} = \arg\min_{X} \sum_{i=1}^{n} \frac{(\hat{d}_i - \|X - X_i\|)^2}{\sigma_i^2}, \tag{7.46}$$

which is in essence a weighted least-squares solution with weights set to $\beta_i = 1/\sigma_i^2$. However, for a static MS, the variance of the TOA measurements may not be significantly different for LOS and NLOS FRPs. Still, the bias in the NLOS distance measurements may degrade the localization accuracy. Hence, in Guvenc et al. (2007, 2008), an alternative weighting technique was proposed, which uses certain statistics of the multipath components of the received signals. In particular, the kurtosis, mean excess delay and root mean square (RMS) delay spread of the received signal are used to evaluate the likelihood of the received signal to be in LOS. The likelihood values are then used to evaluate the weighting parameters β_i.

7.4.2 Residual-weighting algorithm

The Rwgh algorithm proposed by Chen (1999) is based on the observation that the residual error is typically larger if NLOS FRPs are used when estimating the MS location, where the residual error is defined as

$$r_{\mathrm{es}}(X) = \sum_{i=1}^{n} (\hat{d}_i - \|X - X_i\|)^2. \tag{7.47}$$

Assuming that there are more than three FRPs available, Rwgh estimates the MS location as follows (Chen 1999). (1) Form $n_{\mathrm{cb}} = \sum_{i=3}^{n} \binom{n}{i}$ range measurement combinations, where $\binom{n}{i}$ denotes the total number of combinations of i FRPs selected from a total of n FRPs. Also, let $\{S_k | k = 1, 2, \ldots, n_{\mathrm{cb}}\}$ denote the set of FRPs for the kth combination. (2) For each set of combinations S_k, compute an intermediate LS location estimate as follows:

$$\hat{X}_k = \arg\min_{X} \{r_{\mathrm{es}}(X;\, S_k)\}, \tag{7.48}$$

where $r_{\mathrm{es}}(X;\, S_k)$ is the residual error when only the FRPs in set S_k are used for calculating the MS location. Also, we define the normalized residual

$$\bar{r}_{\mathrm{es}}(\hat{X}_k;\, S_k) = \frac{r_{\mathrm{es}}(\hat{X}_k;\, S_k)}{|S_k|}, \tag{7.49}$$

where $|S_k|$ denotes the number of elements of S_k. (3) Find the final location estimate by weighting the intermediate location estimates with their corresponding normalized residual errors:

$$\hat{X} = \frac{\sum_{k=1}^{n_{\mathrm{cb}}} \hat{X}_k [\bar{r}_{\mathrm{es}}(\hat{X}_k;\, S_k)]^{-1}}{\sum_{k=1}^{n_{\mathrm{cb}}} [\bar{r}_{\mathrm{es}}(\hat{X}_k;\, S_k)]^{-1}}. \tag{7.50}$$

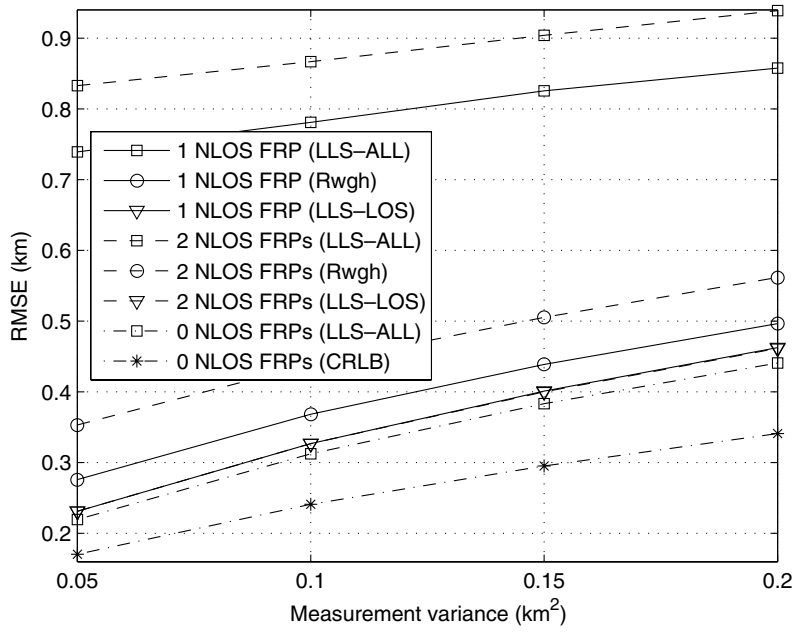


Figure 7.6 Simulations for several different NLOS mitigation techniques.

The main idea behind Rwgh is that it calculates a weighted average of all candidate location estimates, where the location estimates with larger residual errors (i.e., location estimates that are possibly impacted by NLOS propagation) are given smaller weights.

It was shown (Chen 1999) through simulations that Rwgh performs better than choosing the location estimate with the minimum residual error. A suboptimal version of the Rwgh algorithm that has lower computational complexity was proposed by Li (2006). In the latter, instead of considering all combinations of the FRPs (which may be very large in number if n is large), all combinations with $(n-1)$ FRPs are used to calculate the intermediate location estimates and the corresponding residuals. Then, out of n different combinations, the FRP which is not employed in the best estimator (i.e., the estimator corresponding to the combination with the smallest residual error) is discarded. The process is iterated until a predetermined stopping rule is reached (such as when a minimum number of FRPs is reached or when the change in the residual error is small).

Example 7.3 *Consider a cellular localization scenario with parameters as specified in Example 7.2. We have evaluated three different location estimators. The first one was a linear least-squares (LLS-ALL) estimator which uses all seven FRPs for location estimation. A second linear least-squares estimator assumed that the NLOS FRPs were accurately identified and used only the LOS FRPs for estimating the MS location (LLS-LOS). The third estimator was the Rwgh algorithm. For comparison*

purposes, the linear least-squares estimator when all the FRPs are in LOS, and the CRLB in LOS scenarios have also been plotted (Figure 7.6).

Two different NLOS settings have been considered: the first one assumes that only FRP-1 is in NLOS, while the second assumes that both FRP-1 and FRP-2 are in NLOS. The simulation results in Figure 7.6 show that, if the NLOS propagation effects are not handled, there is a significant degradation in the localization performance, as illustrated by the LLS-ALL results. If the NLOS FRPs can be accurately identified, those FRPs can be removed from the localization process to improve the accuracy (LLS-LOS). Since accurate identification of the NLOS FRPs may not always be possible, utilizing the Rwgh technique yields a reasonable accuracy improvement compared with LLS-ALL. Note that if the statistics of the NLOS bias are available, it may be possible to obtain accuracy levels closer to the CRLB, for example by utilizing techniques such as ML-based location estimation.

7.5 Constraint-based Techniques for NLOS Localization

In this section, a different class of NLOS mitigation algorithms that utilize some constraints related to the NLOS measurements will be briefly reviewed.

7.5.1 Constrained LS algorithm and quadratic programming

The two-step ML algorithm discussed by Chan et al. (2006a) is an accurate localization technique in LOS settings; however, it is not robust against NLOS. In Wang et al. (2003), a quadratic programming (QP) technique for NLOS environments was developed. First, Equation (7.8) is written in matrix form as

$$A_1 = \begin{bmatrix} x_1 & y_1 & -0.5 \\ x_2 & y_2 & -0.5 \\ \vdots & \vdots & \\ x_n & y_n & -0.5 \end{bmatrix} \tag{7.51}$$

and

$$\boldsymbol{\theta} = \begin{bmatrix} x \\ y \\ s \end{bmatrix}, \quad \mathbf{p}_1 = \begin{bmatrix} k_1 - \hat{d}_1^2 \\ k_2 - \hat{d}_2^2 \\ \vdots \\ k_n - \hat{d}_n^2 \end{bmatrix}, \tag{7.52}$$

with $s = x^2 + y^2$ being a part of the vector of unknown variables. Then, the mathematical programming is formulated as follows:

$$\hat{\boldsymbol{\theta}}_{cw} = \arg\min_{\boldsymbol{\theta}} (A_1\boldsymbol{\theta} - \mathbf{p}_1)^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (A_1\boldsymbol{\theta} - \mathbf{p}_1), \tag{7.53}$$

$$\text{such that } A_1\boldsymbol{\theta} \leq \frac{1}{2}\mathbf{p}_1, \tag{7.54}$$

where $\Psi = BQB$, and where $B = \text{diag}(2d_1, \ldots, 2d_n)$ and Q is as defined in Equation (7.6). Note that Equations (7.53) and (7.54) constitute a constrained least squares (CLS) algorithm that can be solved using quadratic programming techniques. The intuitive explanation of the CLS is that Equation (7.53) finds a weighted least squares (WLS) solution to the MS location, while the constraint in Equation (7.54) relaxes the equality (which holds in LOS scenarios) into an inequality for the NLOS scenarios. In Wang et al. (2003), a further refining stage is also introduced to incorporate the dependency between s and X.

7.5.2 Linear programming

In Venkatesh and Buehrer (2006, 2007), a linear programming approach was introduced which assumes perfect a priori identification of LOS and NLOS FRPs. As opposed to the identify and discard (IAD) type of algorithms (see Section 7.7), it does not discard NLOS FRPs, but uses them to construct a linear feasible region for the MS location. The location estimate is obtained using a linear programming technique that employs only the LOS FRPs, with the constraint that it has to be within the feasible region obtained.

If the ith FRP is identified to be in NLOS, we have the nonlinear constraint

$$\|X - X_i\| \leq \hat{d}_i, \tag{7.55}$$

which is used in Wang et al. (2003) to formulate a quadratic programming technique for the two-step ML algorithm (without the assumption that the NLOS FRPs are accurately identified). In order to linearize the constraints in Equation (7.55), we can relax them to (Larsson 2004)

$$x - x_i \leq \hat{d}_i, \quad -x + x_i \leq \hat{d}_i, \tag{7.56}$$

$$y - y_i \leq \hat{d}_i, \quad -y + y_i \leq \hat{d}_i, \tag{7.57}$$

where $i = 1, 2, \ldots, n$. In essence, the above is equivalent to relaxing the circular constraints into rectangular (i.e., square) constraints for the purpose of linearizing the equations (see Figure 7.7). By defining some slack variables,[4] Equations (7.56) and (7.57) can be converted to equalities that define the feasible region. Then, the MS location is estimated by using only the LOS FRPs for minimizing an objective function, with the constraint that the location estimate should be within the feasible region obtained using both the NLOS and LOS FRPs.

7.5.3 Geometry-constrained location estimation

In Chen and Feng (2005b), a geometry-constrained location estimation method was proposed, which uses the two-step ML technique of Chan and Ho (1994) with some

[4] A slack variable is a variable which is used to turn an inequality into an equality (e.g., $x < 5 \Rightarrow x + s = 5$).

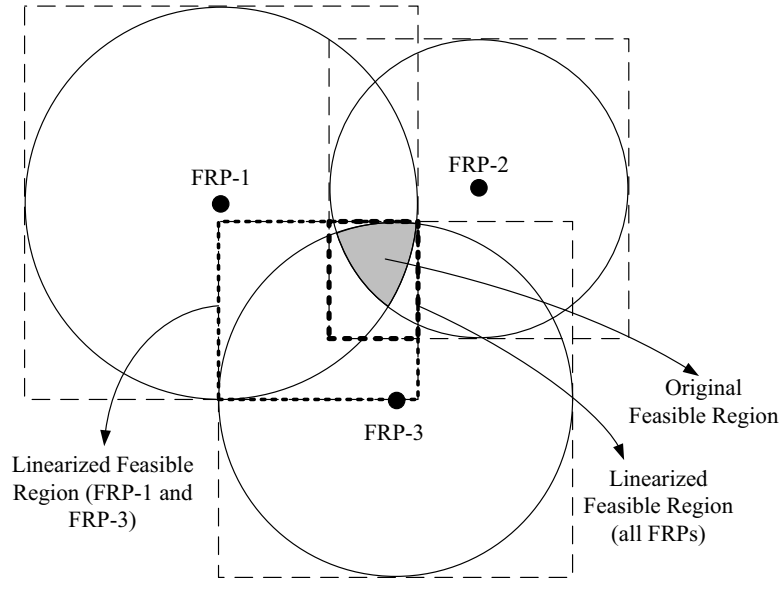


Figure 7.7 Illustration of the constrained linear LS technique. The NLOS FRPs are used to determine the feasible region. For the original (nonlinear) model, the feasible region is obtained from the intersections of the circles. For the linear model, the feasible region is obtained from the intersections of the squares (Guvenc and Chong 2009).

additional parameters to incorporate the geometry of the FRPs (only a scenario with three FRPs was considered). Let the intersection points of the three circles in Equation (7.8) be $X_A = [x_A, y_A]^{\mathrm{T}}$, $X_B = [x_B, y_B]^{\mathrm{T}}$ and $X_C = [x_C, y_C]^{\mathrm{T}}$. Then, a constrained cost function, which is referred to as the *virtual distance*, is defined as (Chen and Feng 2005b)

$$\gamma = \sqrt{\frac{1}{3}[\|X - X_A\|^2 + \|X - X_B\|^2 + \|X - X_C\|^2]}. \tag{7.58}$$

For an expected position X_{e} of the MS, we can calculate the *expected virtual distance* using Equation (7.58) as $\gamma_{\mathrm{e}} = \gamma + n_{\mathrm{e}}$, with n_{e} denoting the noise in the expected virtual distance. The coordinates for X_{e} are chosen as (Chen and Feng 2005b)

$$X_{\mathrm{e}} = \omega_1 X_A + \omega_2 X_B + \omega_3 X_C, \tag{7.59}$$

where the weights are obtained as

$$\omega_i = \frac{\sigma_i^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}, \quad i = 1, 2, 3. \tag{7.60}$$

Basically, it is assumed that for NLOS FRPs, the measurement variance will be larger and the weights defined in Equation (7.60) will move X_{e} towards the center of the NLOS FRP circle, yielding a better location estimate.

These geometric constraints are then incorporated into the two-step ML algorithm of Chan and Ho (1994) by updating A_1 and $\mathbf{p}_1$ in Equations (7.51) and (7.52) as follows (Chen and Feng 2005b):

$$\tilde{A}_1 = \begin{bmatrix} & A_1 & \\ \gamma_x & \gamma_y & -0.5 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{p}}_1 = \begin{bmatrix} \mathbf{p}_1 \\ \gamma_k - \gamma_e^2 \end{bmatrix}, \tag{7.61}$$

where

$$\gamma_x = \frac{1}{3}(x_A + x_B + x_C), \tag{7.62}$$

$$\gamma_y = \frac{1}{3}(y_A + y_B + y_C) \tag{7.63}$$

and

$$\gamma_k = \frac{1}{3}(x_A^2 + x_B^2 + x_C^2 + y_A^2 + y_B^2 + y_C^2). \tag{7.64}$$

The geometric constraints are also incorporated into other variables as

$$\tilde{B} = \text{diag}(d_1, d_2, d_3, \gamma) \tag{7.65}$$

and

$$\tilde{\eta} = [\eta_1, \eta_2, \eta_3, \eta_\gamma]^{\text{T}}. \tag{7.66}$$

Then, the two-step ML algorithm given in Chan and Ho (1994) is employed to solve for the MS location using the variables updated with the geometric constraints, thus improving the performance in NLOS scenarios.

7.5.4 Interior-point optimization

In Kim et al. (2006), an interior-point optimization (IPO) method was proposed for finding the optimum location estimate in the presence of NLOS bias. By linearizing the system using a Taylor series approximation, the nonlinear function $D(X)$ in Equation (7.3) can be linearized around a reference point X_0. Then, neglecting the higher-order terms, we can write

$$D(X) \approx D(X_0) + H_0(X - X_0), \tag{7.67}$$

where H_0 is the Jacobian matrix of $D(X)$ around X_0, which is defined as

$$H_0 = \begin{bmatrix} \dfrac{\partial d_1}{\partial x} & \dfrac{\partial d_2}{\partial x} & \cdots & \dfrac{\partial d_n}{\partial x} \\ \dfrac{\partial d_1}{\partial y} & \dfrac{\partial d_2}{\partial y} & \cdots & \dfrac{\partial d_n}{\partial y} \end{bmatrix}^{\text{T}}_{X=X_0}. \tag{7.68}$$

Note that the reference point X_0 must be chosen sufficiently close to the true location in order for Equation (7.67) to be valid. By substituting Equation (7.67)

into Equation (7.46) and by means of some manipulations, we may obtain a linear system which can be written in a matrix form and solved using a linear LS estimator. A more accurate iterative technique could use this LS estimate as an intermediate estimate, plug it into Equation (7.67) to relinearize the system around it and iterate until convergence (Caffery and Stüber 1998).

By using Equation (7.67), a linearized measurement vector is defined as follows (Kim et al. 2006):

$$Z = H_0 X + B + \eta. \tag{7.69}$$

Then, if NLOS bias is neglected, the bias-free position estimate is given by (see Section 5.2.1)

$$\hat{X} = (H_0^{\mathrm{T}} Q^{-1} H_0)^{-1} H_0^{\mathrm{T}} Q^{-1} Z. \tag{7.70}$$

On the other hand, if the bias vector B is known, a more accurate bias-free location estimate can be derived (Kim et al. 2006):

$$\hat{X}^* = \hat{X} + VB, \tag{7.71}$$

where

$$V = -(H_0^{\mathrm{T}} Q^{-1} H_0)^{-1} H_0^{\mathrm{T}} Q^{-1} \tag{7.72}$$

is a bias correction matrix. However, in reality, B is unknown and has to be estimated.

In order to estimate B from Equation (7.69), the observed bias metric is defined as (Kim et al. 2006)

$$Y = Z - H_0 \tilde{X}, \tag{7.73}$$

which can be simplified to $Y = SB + W$, where $S = I + H_0 V$ and the *bias noise* is given by

$$W = H_0(X - \hat{X}) - \eta. \tag{7.74}$$

Then, the following constrained optimization problem is defined to estimate the NLOS bias errors (Kim et al. 2006):

$$\hat{B} = \arg\min_B (Y - SB)^{\mathrm{T}} Q_w^{-1} (Y - SB), \tag{7.75}$$

$$\text{such that} \quad b_i \in B_i, \quad i = 1, 2, \ldots, n, \tag{7.76}$$

where $B_i = [l_i, u_i]$ is the a priori information for the range of b_i lower-bounded by $l_i \geq 0$ and upper-bounded by u_i, and Q_w is the covariance matrix of W.

In order to solve the constrained optimization problem in Equations (7.75) and (7.76), an IPO technique was used in Kim et al. (2006). In particular, Equations (7.75) and (7.76) were modified to

$$\hat{B} = \arg\min_B (Y - SB)^{\mathrm{T}} Q_w^{-1} (Y - SB), \tag{7.77}$$

$$\text{such that} \quad g_i(b_i) - s_i = 0 \text{ and } s_i > 0, \ i = 1, \ldots, n, \tag{7.78}$$

where s_i is a slack variable and $g_i(b_i)$ is a barrier function that satisfies $g_i(b_i) > 0, \ \forall b_i \in [l_i, u_i]$. A smooth second-order function that is generally used and which satisfies this requirement is $g_i(b_i) = (u_i - b_i)/(b_i - l_i)$. Then, Equations (7.77) and (7.78) are solved by minimizing the following Lagrangian (Kim et al. 2006):

$$\mathcal{L}(B, \boldsymbol{\lambda}, \mathbf{s}) = (Y - SB)^{\mathrm{T}} Q_w^{-1} (Y - SB) - \mu \sum_{i=1}^{n} \ln s_i - \boldsymbol{\lambda}^{\mathrm{T}} (\mathbf{g}(b) - \mathbf{s}), \tag{7.79}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers and $\mathbf{g}(b)$ and $\mathbf{s}$ are obtained upon stacking $g_i(b_i)$ and s_i into $n \times 1$ vectors. Note that the logarithmic barrier function, with μ as a positive bias parameter

$$\mu \sum_{i=1}^{n} \ln s_i, \tag{7.80}$$

ensures that $s_i > 0$ and the bias error is always within $[l_i, u_i]$.

The solution to Equation (7.79) can be obtained by differentiating Equation (7.79) with respect to B, $\boldsymbol{\lambda}$ and $\mathbf{s}$, and solving the resulting equations together to obtain B. Once an estimate of the bias vector B is obtained, the bias correction matrix in Equation (7.72) is used to calculate the bias-free location using Equation (7.71).[5] The simulation results reported in Kim et al. (2006) show that better accuracies can be obtained with IPO than with Rwgh or iterative LS algorithms in NLOS scenarios.

7.6 Robust Estimators for NLOS Localization

Robust estimators are commonly used to suppress the impact of outliers in a given set of data, and several different classes of robust estimators have already been used in the literature for NLOS mitigation purposes. Below, a few of the popular robust estimators considered for NLOS mitigation are briefly reviewed.

7.6.1 Huber *M*-estimator

The *M*-estimators, which are of the "ML type" of estimators, are one class of robust estimators that have been considered for NLOS mitigation purposes. As discussed in the previous sections, the ML algorithm tries to maximize a function of the form

$$\prod_{i=1}^{n} \mathrm{f}(x_i), \tag{7.81}$$

which is equivalent to minimizing $\sum_{i=1}^{n} - \log \mathrm{f}(x_i)$. In the presence of outliers,[6] the ML algorithm fails to yield accurate results. A generalized form of the ML algorithm referred to as the *M*-estimator, which was introduced in Huber (1964),

[5] This is similar to the LOS construction algorithm discussed in Caffery (1999).

[6] An outlying observation, or outlier, is one observation that appears to deviate markedly from other members of the sample in which it occurs (Grubbs 1969).

aims at minimizing

$$\sum_{i=1}^{n} \rho(x_i), \tag{7.82}$$

where $\rho(.)$ is a convex function. For the Huber M-estimator, $\rho(.)$ is defined as (Petrus 1999)

$$\rho(\nu) = \begin{cases} \nu^2/2 & |\nu| \leq \xi, \\ \xi|\nu| - \xi^2/2 & |\nu| > \xi, \end{cases} \tag{7.83}$$

which is not strictly convex and, therefore, minimization of the objective function yields multiple solutions (Petrus 1999).

In Sun and Guo (2004), the M-estimator, used to estimate the MS location in the presence of NLOS bias, was defined as

$$\hat{X} = \arg\min_{X} \left\{ \sum_{i=1}^{n} \rho((\hat{d}_i - \|X - X_i\|)/\sigma_i) \right\}. \tag{7.84}$$

Simulation results (Sun and Guo 2004) show that the M-estimator outperforms the conventional LS estimator, especially for large NLOS bias errors. If the bootstrapping technique[7] is used in conjunction with the M-estimator, the accuracy can be improved even further (Sun and Guo 2004).

7.6.2 Least median squares

In Li et al. (2005), a least-median-of-squares (LMS) technique was proposed for NLOS mitigation; this is one of the most commonly used robust fitting algorithms. It can tolerate a proportion of outliers as high as 50% in the absence of noise. The location estimate of the MS using the LMS solution is given by (Li et al. 2005)

$$\hat{X} = \arg\min_{X} \{\text{med}_i(\hat{d}_i - \|X - X_i\|)^2\}, \tag{7.85}$$

where $\text{med}_i(\theta(i))$ is the median of $\theta(i)$ over all possible values of i. Since calculation of Equation (7.85) is computationally intensive, Li et al. (2005) proposed a lower-complexity implementation that uses random subsets of X_i to obtain several candidate $\hat{X}$, and the one with the least median of residue is selected as the solution. It should be noted that Li et al. (2005) uses the LMS algorithm in the context of security, where some of the measurements may be outliers in the presence of certain attacks against the localization system. A parallel approach that uses the LMS algorithm for NLOS mitigation purposes has been reported in Casas et al. (2006).

[7] Bootstrapping can be defined as obtaining the statistics of an estimator by measuring those statistics by sampling from an approximating distribution, such as the empirical distribution of the observed data (Chernick 1999).

7.6.3 Other robust estimation options

Besides the M-estimation and the LMS techniques discussed above, there are some other robust estimation techniques (Ripley 2004) which, to the best of our knowledge, have not been considered in detail for NLOS mitigation purposes. For example, the least-trimmed-squares (LTS) technique aims at minimizing the sum of squares for the smallest n of the residuals. On the other hand, the S-estimator, which attempts to find a line minimizing a robust estimate of the scale of the residuals, is quite robust to outliers; however, it is not as efficient as the M-estimator. Finally, the MM-estimator can be considered as an M-estimator which starts at the coefficients given by the S-estimator (Knight and Wang 2009). With more computational complexity compared with the other two techniques, the MM-estimator has both good robustness and good efficiency.

There are also some other parametric approaches for robust regression, where the normal distribution is replaced with a heavy-tailed distribution in order to model the outliers. For example, the t-distribution can be a good distribution for modeling the outliers in a Bayesian robust regression algorithm (Taylor 2002). Alternatively, a mixture of zero-mean Gaussian distributions (i.e., a contaminated normal distribution) with different variances can be considered to model the noise in the presence of NLOS bias:

$$\eta_i \sim (1-\epsilon)\text{Norm}(0, \sigma^2) + \epsilon\, \text{Norm}(0, c_{\text{nlos}}\sigma^2), \tag{7.86}$$

where the second Gaussian distribution is intended to capture the outliers, and $\epsilon < 1$ is a small number (typically smaller than 0.1) that characterizes the impact of the outliers together with the constant term c_{nlos}.

7.7 Identify and Discard Techniques for NLOS Localization

As discussed in the beginning of Section 7.3, one of the simplest techniques to mitigate NLOS effects is to identify the NLOS FRPs and discard them during localization (i.e., to find the MS location using only the LOS FRPs). In fact, the ML estimator of Qi et al. (2006) illustrated in Figure 7.5(a) and the ML estimator of Riba and Urruela (2004) discussed in Section 7.3.1 are also estimators of the identify and discard (IAD) type. Moreover, the WLS technique discussed in Section 7.4.1 also becomes an IAD type of technique if the weights are set to 0 for NLOS FRPs and 1 for LOS FRPs. A common problem in all of these approaches is accurate identification of the NLOS FRPs. In this section, we will review another IAD-based technique, which uses a residual test algorithm to identify the NLOS FRPs.

7.7.1 Residual test algorithm

The residual test (RT) algorithm proposed in Chan et al. (2006b) falls into the group of algorithms where the MS is localized using only the LOS FRPs. Hence, the NLOS

FRPs have to be correctly identified and discarded, which is achieved as follows. First, by using the approximate ML algorithm discussed in Chan et al. (2006a) and by employing different combinations of FRPs,[8] different location estimates $\hat{X}_m$ are computed, and we define

$$S_0 = \sum_{i=3}^{n} \binom{n}{i}. \tag{7.87}$$

For each m, the squares of the normalized residuals[9] are computed as

$$\chi_x^2(m) = \frac{[\hat{x}_m - \hat{x}_{S_0}]^2}{I_x(m)} \quad \text{and} \quad \chi_y^2(m) = \frac{[\hat{y}_m - \hat{y}_{S_0}]^2}{I_y(m)}, \tag{7.88}$$

where $I_x(m)$ and $I_y(m)$ are obtained from Equation (7.17) and are the CRLBs[10] for x and y dimensions and for the mth hypothesis. Without loss of generality, S_0 indexes the case when all the FRPs are used in localization.

If all the FRPs in the mth hypothesis are in LOS, we have $\chi_x(m) \sim \text{Norm}(0, 1)$ and $\chi_y(m) \sim \text{Norm}(0, 1)$. Hence, $\chi_x^2(m)$ and $\chi_y^2(m)$ have centralized chi-square distributions with one degree of freedom. On the other hand, if there is at least one NLOS FRP in the mth hypothesis, both random variables have noncentralized chi-square distributions with noncentrality parameters depending on the NLOS bias. This observation suggests that if the PDFs of $\chi_x^2(m)$ and $\chi_y^2(m)$ (which should ideally be the same) can be identified correctly, it allows one to determine whether all the FRPs are in LOS or not. This can be simply achieved by a threshold test. For example, for a case with seven FRPs as in Chan et al. (2006b), an appropriate threshold to characterize the chi-square PDF for the LOS case has been determined to be 2.71. If the area under the PDF to the right of this threshold is larger than 0.1, the random variable is identified as having a noncentralized chi-square distribution. If the PDF is determined to be a noncentralized chi-square distribution, this implies that there is at least one NLOS FRP. Then, the algorithm forms $\binom{n}{n-1} = n$ sets of FRPs, with $(n - 1)$ FRPs in each set. For each of these n sets,

$$\sum_{i=1}^{n-1} \binom{n-1}{i} \tag{7.89}$$

estimates of X_m are obtained. If any of these sets is found to be distributed according to a centralized chi-square distribution using the threshold test, then the number of LOS FRPs is $(n - 1)$, and the FRPs within that particular set are used to estimate the MS location. Otherwise, the algorithm iterates until there are at least three LOS

[8]The summation in Equation (7.87) starts from 3 since at least three FRPs are required for location estimation in 2D.

[9]Note that the definition of "residual" in the RT algorithm is different from that in the Rwgh algorithm.

[10]The computation of the CRLB requires the MS location X. Since that is not available, X_{S_0} is used as an approximation for X to compute an approximate CRLB.

Table 7.1 Overview of TOA-based localization algorithms for use in LOS and NLOS scenarios

	Algorithm name	References	Summary	Complexity, a priori knowledge etc.
ML-type algorithms	ML algorithm utilizing NLOS statistics	Gezici and Sahinoglu (2004); Qi et al. (2006)	In Gezici and Sahinoglu (2004), the X that maximizes the joint probability density function of the observations in NLOS scenarios is selected. A MAP estimator utilizing the NLOS bias statistics is introduced in Qi et al. (2006).	The probability density function of the NLOS bias and the distance measurements are assumed known, and the method requires a search over possible MS locations.
	IAD-based ML algorithm	Qi et al. (2006); Riba and Urruela (2004)	Uses the ML principle to discard the NLOS FRPs. Then, only the LOS FRPs are used in location estimation.	Need to collect n_{trn} TOA measurements for each FRP to capture the noise statistics (Riba and Urruela 2004). There may always be a possibility of misidentification of the LOS FRPs.
LS algorithms	Weighted LS	Guvenc et al. (2007, 2008)	Uses some appropriate weights (e.g., using the variance of the distance measurements or the statistics of the multipath components) to assign less reliability to NLOS FRPs.	For a static MS, the variance information may not be very different for LOS and NLOS FRPs.
	Residual-weighting algorithm	Chen (1999)	Different possible combinations of FRPs are considered. Then, the corresponding location estimates are weighted with the inverses of the residual errors to obtain the final location estimate.	Need to solve for $n_{\text{cb}} = \sum_{i=3}^{n} \binom{n}{i}$ location estimates for different hypotheses before weighting them.
Constrained localization techniques	Constrained LS with QP	Wang et al. (2003)	A two-step ML technique is used to obtain an estimate of the MS, with a quadratic constraint as given in Equation (7.54).	May have high computational complexity.
	Constrained LS with LP	Venkatesh and Buehrer (2006, 2007)	The NLOS FRPs are used to obtain a feasible region composed of squares. Then, the LOS FRPs are used to solve for the MS location via the LLS-1 technique so that the solution is within the feasible region.	Linear constraints yield a less complex (but a coarser) solution compared with the quadratic constraints.
	Geometry-constrained localization	Chen and Feng (2005b)	A constraint related to the intersection points of circles is incorporated into the two-step ML algorithm.	Slightly more complex than the two-step ML algorithm.
	Interior-point optimization	Kim et al. (2006)	First, the NLOS bias values are estimated with IPO. Then, the NLOS bias estimates are used in a WLS solution (linearized using a Taylor series approximation).	Bias estimation through IPO may be computationally complex.
Robust estimators	M-estimators	Sun and Guo (2004)	Employs a convex function $\rho(v)$ to capture the effects of NLOS bias values in an "ML-type" of estimator.	Need to tune $\rho(v)$ appropriately. Better alternatives such as S-estimators (more robust) and MM-estimators (both robust and efficient) are available.
	Least median of squares	Li et al. (2005)	The location that minimizes the LMS of the residual is selected as a location estimate.	Robust up to 50% of outliers. More computationally complex than the NLS method.
Identify and discard techniques	Residual test algorithm	Chan et al. (2006b)	Identifies and discards the NLOS FRPs. The residual errors are normalized by the CRLBs, and the resulting variables are checked to find if they have a centralized or noncentralized chi-square distribution.	Computationally complex owing to testing numerous hypotheses, Delta test etc.

FRPs. Since three FRPs do not have a sufficient number of realizations for a reliable application of the RT algorithm, the delta test procedure has been proposed, which takes two FRPs first and then combines them with one of the rest of the FRPs to check if all three of those FRPs are in LOS. Simulation results show that the proposed technique outperforms the Rwgh (Chen 1999) and CLS (Wang et al. 2003) algorithms, and can achieve the CRLB if the number of LOS FRPs is larger than half of the total number of FRPs.

7.8 Conclusions

In this chapter, an extensive survey of TOA-based localization and NLOS mitigation techniques was presented. While some techniques discussed in this chapter can perform close to the CRLB, they may require high computational complexity and the availability of various types of prior information. For example, prior information regarding the NLOS bias statistics may not be available in many scenarios, forcing the use of techniques such as the residual-weighting algorithm. In Table 7.1, we provide a brief summary of the various NLOS mitigation methods and of their complexity and requirements.

Several of the NLOS mitigation techniques reviewed may also be extended to TDOA-, AOA- and RSS-based systems after some simple modifications. For example, robust estimation techniques similar to those discussed in Section 7.6 may also be utilized for handling outliers in TDOA, AOA or RSS measurements. Using techniques similar to those discussed in Section 7.4, appropriate weights may be used to suppress TDOA, AOA or RSS measurements that are subject to NLOS bias. Extensions of the maximum-likelihood and constrained-based techniques as discussed in Sections 7.3 and 7.5 to other localization scenarios are also possible. All in all, the achievement of practical, efficient, low-complexity localization techniques in the presence of NLOS propagation still requires further research. The authors hope that this chapter will serve as a valuable resource for researchers for evaluating the merits and trade-offs of the various techniques available in the present state of the art with the aim of developing more efficient and practical NLOS mitigation methods.

8

Positioning Systems and Technologies

Andreas Waadt, Guido H. Bruck and Peter Jung

Lehrstuhl für KommunikationsTechnik, Universität Duisburg-Essen, Germany

8.1 Introduction

Applications requiring positioning in mobile networks have gained importance in recent years. Examples include not only LBSs, but also enhanced emergency call services such as the North American E911 and the newly defined Emergency Call (eCall), which is planned to become a European standard as part of the eSafety initiative of the European commission (European Commission 2005). Mobile network providers already offer LBSs along with positioning, although the accuracy and reliability of the present techniques do not yet meet the requirements of such services in all cases.

One of the most popular positioning systems in navigation is the GPS, which is satellite based, widely available and quite accurate, as long as the receiver to be localized and the satellites have an LOS connection. Satellite positioning systems are discussed in Section 8.2.

Since most common mobile phones do not support the GPS, network providers evaluate the position of the network's base stations (BSs) which have connections to the mobile stations (MSs) to be located. The accuracy of the present position estimation techniques can be increased by evaluating additional protocol data from the cellular network, for example in the case of GSM. Some additional information

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

which can be evaluated is measured by the mobile station and available only there. This leads to the approach of mobile-assisted position estimation techniques (Waadt et al. 2008). Section 8.3 focuses on positioning in cellular networks.

With less area-wide coverage but nevertheless with interesting positioning capabilities and applications, wireless local area networks (WLANs) are increasingly prospering. Examples include Wi-Fi and UWB systems, the latter allowing precise real-time localization and tracking (see Section 8.4). Positioning is also discussed in that section in the context of dedicated systems, where the solutions include infrared, RFID and ultrasound systems.

Positioning in ad hoc networks has been considered an important asset for their performance improvement. This topic is briefly addressed in Section 8.5.

Finally, hybrid positioning solutions, which combine two or more different communication concepts, are being considered in order to facilitate improved ease of use in terms of low latency of position estimation. At present the combination of cellular and WLAN positioning and also A-GNSS, e.g., A-GPS, is of interest. These concepts will be concisely presented in Section 8.6. Section 8.7 concludes this chapter.

8.2 Satellite Positioning

8.2.1 Overview

Positioning methods are generally based on the determination of distances, directions or angles between known points and unknown points. In geodesy (from the Greek *geodaisia*, division of the Earth), which is the science of measuring the Earth and primarily concerns positioning, the known points are often trigonometric points, whose locations have been determined in the past. The first and most fundamental reference point of a geodetic coordinate system is the geodetic reference datum. Together with the reference ellipsoid, a mathematical approximation of the Earth's shape, the geodetic datum defines the coordinate system. A popular reference ellipsoid is the European Bessel Ellipsoid, which is often used in combination with the Potsdam datum as a reference datum; in contrast, the World Geodetic System 1984 (WGS84) uses a reference frame of 12 fundamental reference points, called the geodetic data, which are distributed over the Earth. The principle of using trigonometric points as references for positioning requires a broad and reliable network of known trigonometric points, which can be seen from every point whose location could potentially be determined. Our ancestors used known stars and constellations as a reference for navigation. One essential advantage of this archaic method is that it provides global coverage, since the stars can be seen from nearly every location on Earth, as long as there is a clear view of the night sky.

Nowadays, known stars and constellations have been replaced by the satellites of Global Navigation Satellite Systems (GNSSs). For positioning, it is sufficient to receive the radio signals of at least four satellites whose positions are known.

Mentionable GNSSs include the American Transit and GPS, the Russian Globalnaya Navigationnaya Sputnikovaya Sistema (GLONASS), the European GALILEO and the Chinese Compass. The basic principles and the mathematical treatment will be given in Section 8.2.2.

A historical review of GNSSs forms part of Section 8.2.3. Satellite-Based Augmentation Systems (SBASs) also include stationary transmitters, which transmit correction signals as to improve the accuracy of the positioning. Some examples of SBASs are the American Wide Area Augmentation System (WAAS), the Japanese Multifunctional Satellite Augmentation System (MSAS), the Indian GPS-Aided Geo-Augmented Navigation system (GAGAN) and the European Geostationary Navigation Overlay Service (EGNOS). Finally, the future of GNSSs, in particular GPS III and GALILEO, will be touched upon.

Accuracy and reliability are considered in Section 8.2.4, leading to the identification of the drawbacks of satellite positioning when applied to mobile positioning (Section 8.2.5).

8.2.2 Basic principles

GNSSs make use of satellites which broadcast their positions and clock times. To estimate its location, a receiver must receive the signals of at least four satellites at the same time. Then, it calculates the time of flight and, from that, its distance from each satellite. With such distances and the position of the satellites, the location of the receiver can be calculated, including longitude, latitude and altitude.

The basic principle of the localization procedure, previously described in Section 4.4.2.1, is also illustrated by the following example. The measured time of flight and the resulting calculated distance to a satellite define a sphere around the satellite itself (Figure 8.1). In the case of error-free measurements and calculations, the receiver must be located on the surface of this sphere. The measurement of the signal of a second satellite defines a second sphere. The two spheres intersect in a circle, limiting the possible positions where the receiver may be located. When a third sphere is evaluated by exploiting the signal from a third satellite, the three spheres will intersect in two points, one at the location of the receiver and one at an ambiguous point in distant space. This solution should actually be sufficient to determine the position of the receiver when it is known a priori that the receiver is located on the Earth. However, the time-of-flight measurements require that the satellites and the receiver use accurately synchronized clocks. The satellites do in fact use accurately synchronized atomic clocks, but the clock of a receiver may not be sufficiently synchronized to allow accurate time-of-flight measurements. This would result in enormous positioning errors, if not compensated. The unknown time difference between the receiver's clock and the satellites' clock is a fourth unknown variable to be determined, requiring a fourth equation and therefore a fourth satellite, which must be seen by the receiver. The localization of a receiver requires therefore at least four satellites, assuming that the positions of the satellites are accurately known.

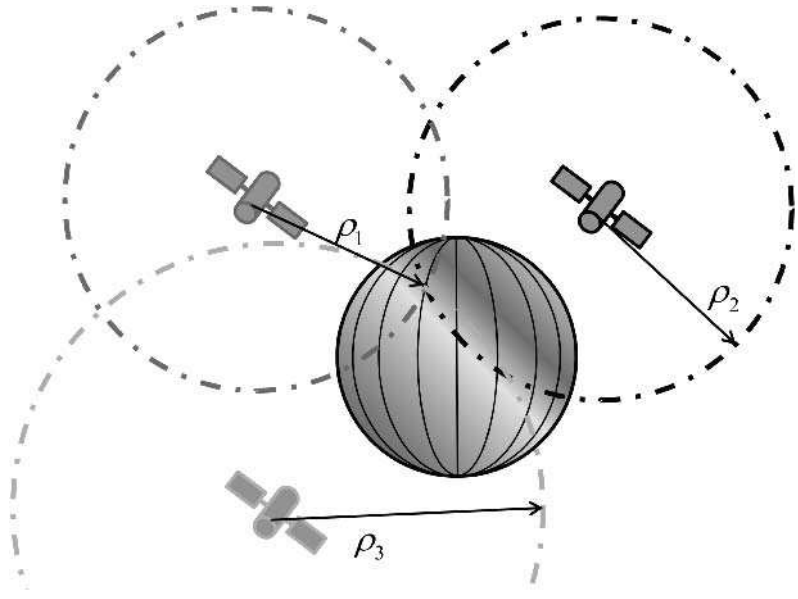


Figure 8.1 Distance measurements to satellites define orbits intersecting in a point.

In practice, GNSS receivers usually evaluate signals from more than four transmitters in order to minimize the drawback of error-prone distance measurements to the satellites. The reasons for measurement errors include changes of the travel time of the radio signal in the ionosphere (see Section 4.5.1.6), and noise and quantization errors in the receiver (Xu 2003). Owing to these inaccuracies, the spheres resulting from the measurements of the satellites' signals will generally not intersect in one or two locations, and the equation system will not result in a unique solution. The receiver's position must consequently be estimated, for instance with WLS estimators or Kalman filtering (see Chapter 5).

8.2.2.1 Mathematical background

Let us assume that a receiver receives n signals from n satellites and extracts timestamps when a signal is received. The signal transmitted by the ith satellite contains the timestamp t_i^{tx}, $i \in \{1, \ldots, n\}$, which is the point in time of transmission from the satellite. The clocks of the n satellites are assumed to be accurately synchronized, and all timestamps provided by the satellites are in terms of the same satellite reference time. In the satellite reference time, the signal reaches the receiver at a point in time t_i. However, since the receiver is not perfectly synchronized with the satellites, the timestamp

$$\tilde{t}_i = t_i + \tau, \quad i \in \{1, \ldots, n\}, \tag{8.1}$$

extracted by the receiver is a pseudo-time or rough estimate of the true time of reception, t_i, plus a deterministic time measurement error, τ, which is the offset between the satellite reference time and the receiver's clock. Stochastic components resulting from propagation conditions are not considered at this point. The receiver can calculate the pseudo-distances

$$\hat{d}_i = (\tilde{t}_i - t_i^{(\mathrm{tx})})c = \underbrace{(t_i - t_i^{(\mathrm{tx})})c}_{d_i} + \underbrace{\tau c}_{b}, \quad i \in \{1, \ldots, n\}, \tag{8.2}$$

where c is the speed of light and $\tilde{t}_i$ are the timestamps. The pseudo-distances $\hat{d}_i$ contain the true distances d_i, equal to $(t_i - t_i^{(\mathrm{tx})})c$, and a bias b, equal to τc, which results from the clock offset in Equation (8.2) and does not depend on i. The coordinates

$$X_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}, \quad i \in \{1, \ldots, n\}, \tag{8.3}$$

of the n satellites are transmitted with the satellites' signals and are therefore known by the receiver. The movement of the satellites and the Earth's rotation during the transmission have not yet been considered. The receiver's coordinates

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{8.4}$$

then need to be determined. The above equations (8.2) and (8.4) contain four unknown variables, namely the receiver's coordinates x, y, z and the bias b. Hence, the calculation of the receiver's position requires measurements of at least four satellites' signals, leading to the equation

$$(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 = d_i^2 = (\hat{d}_i - b)^2, \quad i \in \{1, \ldots, n\}, \tag{8.5}$$

with $n \geq 4$. If $n = 4$, Equation (8.5) yields a four-dimensional system of quadratic equations with four unknown variables and a unique solution. If the receiver sees more than four satellites, the method of least-squares (see Section 5.2) can be used to obtain the optimal solution.

In the following, we illustrate a possible way to solve the above equation. By subtracting the ith equation from the first equation, the quadratic components are eliminated and we obtain the linear equation

$$\kappa_i - \kappa_1 - 2x_i^* x - 2y_i^* y - 2z_i^* z = \hat{d}_i^2 - \hat{d}_1^2 - 2\hat{d}_i^* b, \tag{8.6}$$

where

$$\kappa_i = x_i^2 + y_i^2 + z_i^2 \tag{8.7}$$

and

$$w_i^* = w_i - w_1, \quad w \in \{x, y, z, \hat{d}\}. \tag{8.8}$$

Separating the known terms from those with unknown variables in Equation (8.6), i.e., the receiver's coordinates x, y, z and the bias b, we obtain

$$x_i^* x + y_i^* y + z_i^* z - \hat{d}_i^* b = \frac{1}{2}[\kappa_i - \kappa_1 - (\hat{d}_i^2 - \hat{d}_1^2)], \quad i \in \{1, \ldots, n\}. \tag{8.9}$$

Using the system matrix

$$H = \begin{bmatrix} x_1^* & y_1^* & z_1^* & -\hat{d}_1^* \\ x_2^* & y_2^* & z_2^* & -\hat{d}_2^* \\ \vdots & \vdots & \vdots & \vdots \\ x_n^* & y_n^* & z_n^* & -\hat{d}_n^* \end{bmatrix}, \tag{8.10}$$

and the vectors

$$\mathsf{X} = \begin{bmatrix} X \\ b \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ b \end{bmatrix} \tag{8.11}$$

and

$$C = \frac{1}{2} \begin{bmatrix} \kappa_2 - \kappa_1 - (\hat{d}_2^2 - \hat{d}_1^2) \\ \kappa_3 - \kappa_1 - (\hat{d}_3^2 - \hat{d}_1^2) \\ \vdots \\ \kappa_n - \kappa_1 - (\hat{d}_n^2 - \hat{d}_1^2) \end{bmatrix}, \tag{8.12}$$

the $n - 1$ equations in Equation (8.9) can be written in matrix vector notation:

$$H\mathsf{X} = C. \tag{8.13}$$

When $n > 4$, i.e., the receiver sees more than the minimum of four satellites required. Then, the matrix H is not square anymore and cannot be directly inverted. Rather, a pseudo-inverse $H^\dagger$ needs to be determined, for example the Moore–Penrose pseudo-inverse (Penrose 1955),

$$H^\dagger = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}}. \tag{8.14}$$

In Equation (8.14), H^{T} denotes the transposed version of the matrix H, and $(H^{\mathrm{T}} H)^{-1}$ is the inverted matrix of $(H^{\mathrm{T}} H)$. Applying $H^\dagger$ to Equation (8.13), we obtain

$$\mathsf{X} = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} C. \tag{8.15}$$

An elegant solution of Equation (8.15) can be implemented by starting out from

$$(H^{\mathrm{T}} H)\mathsf{X} = H^{\mathrm{T}} C \tag{8.16}$$

and using the two-iteration-step approach illustrated in Jung (1997), which is based on Cholesky and Schur decomposition.

The offset of the receiver's clock with respect to the satellites' reference clock is actually not the only bias. Other systematic sources of error include effects of the atmosphere, of the Earth's rotation and of Einstein's relativity. The large motion velocities and the gravitational potential differences between the satellites and receiver, for instance, lead to nonnegligible relativistic effects. These include a frequency offset between the signals transmitted by the satellites and the signals observed by the receiver, and path range effects caused by time dilation. For the details of the sources of errors, the reader is referred to Kaplan and Hegarty (2006) and Xu (2003).

8.2.3 Satellite positioning systems

8.2.3.1 Introductory remarks

The first GNSS and the predecessor of the well-known GPS was originally called the Navy Navigation Satellite System (NNSS), later called Transit. Starting in 1958, Transit was developed by the US Navy to provide accurate location information to ballistic missiles and submarines. From 1967 it was also used for civil applications. Its accuracy was between 500 and 15 m. Transit was shut down in 1996.

GLONASS is a satellite navigation system developed by the Russian Space Forces. It was deployed during the Cold War as a competitor to the GPS, starting in 1976.

The Chinese Beidou navigation system is a project for the development of an independent satellite navigation system. It is named after the Big Dipper constellation, which is called *Beidou* in Chinese. It is an experimental system of only four satellites, but the final GNSS with global coverage will include 35 satellites and be called Compass.

8.2.3.2 The Global Positioning System

The GPS, officially the Navigational Satellite Timing and Ranging (NAVSTAR) GPS, is the successor of Transit and probably the best known and most popular satellite positioning system for localization and navigation purposes. Developed from the 1970s by the United States Department of Defense, it replaced Transit in 1985.

A constellation of at least 24 satellites transmits signals with a data rate of 50 bit/s and a frame period of 30 s on two frequencies, the L1 signal at 1.57542 GHz and the L2 signal at 1.2276 GHz (additional signals on other frequencies have been proposed and are intended to be deployed). CDMA is used to separate the signals of different satellites. The good autocorrelation properties of the CDMA codes used allow accurate synchronization and therefore accurate determination of the TOA, i.e., the point in time of the radio reception, which is necessary to estimate the time of flight. GPS uses two codes, a public coarse/acquisition code for civil applications and an encrypted precise code for military use (Parkinson and Spilker 1996).

8.2.3.3 Augmentation systems

For further improvement of the accuracy of GNSSs, there are several augmentation systems in use. Most of them are designed to cooperate with the GPS. Examples include the Differential Global Positioning System (DGPS), the US WAAS, the European EGNOS, the Japanese MSAS and the Indian GAGAN. Augmentation systems provide additional information to the receiver. This may include information about the speed of a vehicle, which can be used advantageously if the positioning system is integrated into a car or an aircraft, for instance. In DGPS, a network of fixed base station receivers is used as a reference for the localization of mobile station receivers. After a comparison of time-of-flight measurements with the expected results, a base station receiver can estimate sources of errors, such as ionospheric delays. Since those error sources are nearly constant within small areas, they can be compensated by estimations made at the mobile receiver side.

8.2.3.4 GPS III and GALILEO

The improvement of GNSSs is an ongoing process. In the GPS, for instance, the satellites of the first generation have been shut down and a third generation, GPS III, is to be deployed (Luba et al. 2005). GPS III will allow improved accuracy and reliability, making it more robust against jamming and spoofing. Although similar in design, the European satellite navigation system GALILEO is, in contrast to the GPS, designed only for civil applications and is without a separately encrypted CDMA channel for military use. Furthermore, it is not under military control. As of today, GALILEO (Antonini et al. 2004) is still under construction and will be compatible with GPS III, leading to better coverage and accuracy if used in combination.

8.2.4 Accuracy and reliability

The accuracy of satellite positioning is usually in the range of few tens of meters; this is the case, for example, for GPS and GLONASS. In reality, the accuracy depends on the weather conditions and the sensitivity of the receiver. When augmentation systems are used, such as the DGPS, for example, the accuracy can be improved to the range of meters or centimeters, depending on the distance from the base station receivers. A disadvantage of satellite positioning is the LOS constraint between the receiver and the satellites, which results in satellite positioning not being fully reliable. For details on the accuracy and reliability of GNSSs, the reader is referred to David (2009).

8.2.5 Drawbacks when applied to mobile positioning

Satellite positioning systems are dedicated positioning systems designed for accurate, global navigation, and the accuracy and global availability of satellite positioning are their major strengths compared with other positioning techniques. Furthermore, satellite positioning supports fast tracking of mobile users. However,

two requirements of satellite positioning systems constrain their usability when applied to mobile positioning: (1) the need for an LOS connection between the satellites and the receiver, and (2) knowledge of the orbital elements, locations and ephemerides of the satellites. The first requirement leads to a limited capability of satellite positioning systems in indoor environments and in the case of bad weather conditions. The second requirement implies that the orbital elements must be downloaded before the first position fix is acquired. Owing to the small data rate in satellite positioning systems, it can take minutes to localize the receiver the first time. As a consequence, the receiver must stay on line in order to keep the orbital elements up to date. This fact, together with the energy management approach that satellite positioning systems implement while in normal operational mode, increases the power consumption, which is a critical parameter in mobile applications. This makes satellite positioning systems only conditionally applicable to mobile positioning.

8.3 Cellular Positioning

8.3.1 Overview

The localization of mobile terminals in cellular networks has been discussed over the past ten years, usually focusing on GSM-based solutions (Drane et al. 1998; Kyammaka and Jobmann 2005). It is particularly noteworthy that the techniques discussed in Kyammaka and Jobmann (2005) influenced the standardization of UMTS. For example, one specification (3GPP 2008a) specifies the positioning methods to be supported in UMTS.

Nowadays, many manufacturers offer numerous models of GPS-enabled cell phones. However, the increased cost of GPS-enabled mobiles, as well as their drawbacks when applied to mobile positioning, is reducing the speed at which they are spreading on the market. Owing to the fact that most of the currently circulating cell phones are still without GPS, the following will focus on nonsatellite-based, or GPS-free, positioning, which is achieved by exploiting information from the cellular network. Generally, we distinguish between mobile-assisted GPS-free positioning and mobile- or network-based GPS-free positioning. In mobile-assisted GPS-free positioning, the mobile calculates its position using signals received from base transceiver stations (BTSs), whereas in network-based GPS-free positioning the position of a mobile is determined on a server in the network. Network-based GPS-free positioning is quite common, whereas mobile-assisted GPS-free positioning has not yet been widely introduced.

One simple network-based GPS-free positioning method is based on the cell coverage, by means of evaluating the cell identification. This scheme is commonly deployed by mobile network operators. The position of a mobile connected to a particular BTS, which is identified by its cell ID, is determined by the location of the base station itself. A more advanced network-based GPS-free positioning method using the cell coverage determines the location of a mobile connected to

a multiantenna BTS by evaluating the center of gravity of the sector that the mobile belongs to, thus using that location as the estimated position of the mobile.

Recently, network-based GPS-free positioning techniques have also been using recursive Bayesian filtering (RBF) based on models of the radio environment, i.e., data from network planning and radio propagation prediction in the vicinity of the mobile, and predictions of the user's mobility. Although these techniques facilitate position estimation accuracies of the order of those achieved with GNSSs, the creation and maintenance of the required radio propagation databases and the estimation of mobility are extremely cumbersome. Since positioning services might have only a minor market importance compared with multimedia services, network operators are not much interested in bearing the burden of creating these databases. More elaborate positioning methods evaluate the TOA and TDOA of radio signals. These methods require reasonably accurate synchronization among the base stations, which is usually cumbersome.

Several cellular positioning techniques have already been commercially deployed for emergency call location and for mobile tracking services from mobile network operators, which allow a mobile to be localized using a Web interface, for instance. The Google Maps Mobile Application also provides GPS-free cellular positioning for navigation purposes (Google 2006).

8.3.2 GSM

8.3.2.1 Cell ID

One simple network-based GPS-free positioning method is cell coverage positioning by means of evaluating the cell ID. This scheme is commonly deployed by mobile network operators. The position of a mobile that is connected to a particular BTS, identified by its cell ID, is determined by the location of the base station itself. This scheme is referred to as "BTS connected." This is a very rough, quite inaccurate localization technique, since it does not take the cell's geometry into account. Owing to their directional characteristics, BTS antennas are most often located at the borders of the cell sectors, and this leads generally to large localization errors when the "BTS connected" method is used. If, besides the cell ID, no additional information is known, then the estimation error can be minimized by estimating the mobile station's location as the center of gravity of the cell. The determination of such a center of gravity requires a knowledge of the cell's geometry.

A more advanced network-based GPS-free positioning method using the cell coverage determines the location of a mobile connected to a BTS sector by first evaluating the center of gravity of the sector and using the location of that as the estimated position of the mobile. A simple but reasonably accurate way of determining a virtual center of gravity is the following: (1) determine the distances between the connected BTS and the BTSs in the neighborhood; (2) select the smallest distance value d; (3) divide the cell into sectors, according to the directivity of the BTS antennas; (4) the virtual center of gravity is in the middle of the sector of the

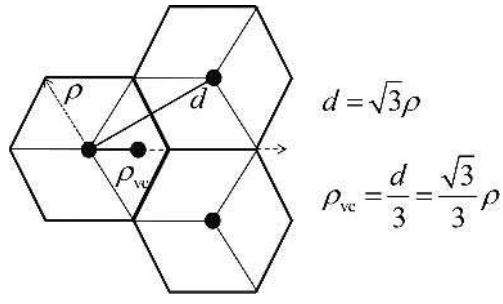


Figure 8.2 Virtual center in a symmetric cellular network.

connected antenna, at a distance $\rho_{vc} = d/3$ from the connected BTS (Figure 8.2). This advanced cell coverage positioning scheme is referred to as the "virtual center" scheme. "BTS connected" and "virtual center" can also be implemented in a mobile-assisted GPS-free positioning fashion.

The aforementioned "virtual center" method assumes a symmetric network geometry, which is usually not the case. An accurate determination of the cell's geometry can be achieved by measurements of the radio signal power from the reference BTS and the neighboring BTSs. This empirical method, though, is cumbersome. The cell geometry can be determined roughly in a simplified way without the need for radio measurements. Let us assume free-space propagation, and an equivalent radiated power (ERP) that is assumed to be the same for all BTSs. In this case, the cell geometry becomes the Voronoi diagram of the BTSs (Baert and Seme 2004).

Figure 8.3 illustrates the construction of the cell geometry for an example where there is a reference BTS surrounded by seven neighboring BTSs. The boundaries of the cells are constructed from the perpendicular bisectors of the line segments between the reference BTS and the neighboring BTSs. If the reference BTS is composed of several antennas with different directional characteristics, then the cell must be divided into appropriate cell sectors. Figure 8.3 shows an example with three cell sectors from three BTSs, one pointing to the north (0°), one pointing to the southeast (120°) and one pointing to the southwest (240°). Assume that the corners of the polygonal cell in Figure 8.3 are numbered from 1 to n and their corresponding positions are given by

$$X_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad i \in \{1, \ldots, (n+1)\}, \quad X_1 = X_{n+1}, \tag{8.17}$$

where x_i and y_i are the coordinates of the ith cell corner in two dimensions and X_{n+1} is equal to X_1. From geometrical theory for the calculation of the area and the centroid of a polygon, it is possible to obtain the area

$$A = \frac{1}{2} \sum_{i=1}^{n} (x_i y_{i+1} - x_{i+1} y_i) \tag{8.18}$$

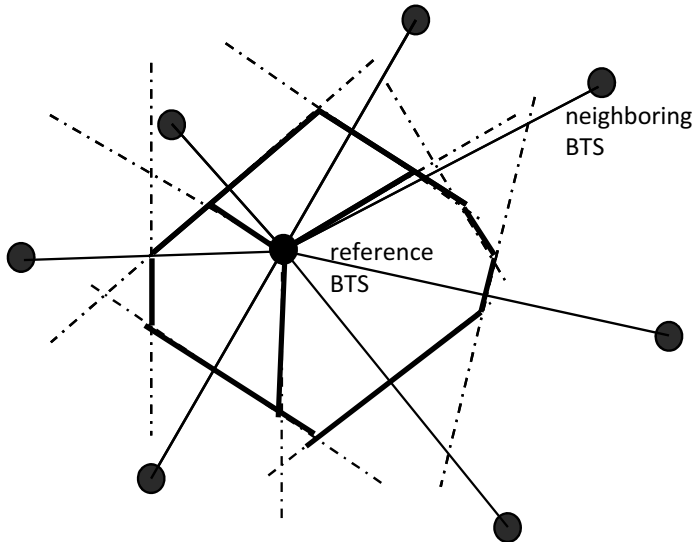


Figure 8.3 Construction of the cell geometry.

and the coordinates of the centroid

$$X = \frac{1}{6A} \begin{bmatrix} \sum_{i=1}^{n} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \\ \sum_{i=1}^{n} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \end{bmatrix}. \tag{8.19}$$

Figure 8.4 shows an example of a computer-constructed cell geometry of a cellular GSM network in Duisburg, Germany. The black marker in the middle of the figure represents the location of a reference BTS. The other six markers near the centers of the three cell sectors represent the centers of gravity of these sectors and the true locations of the test mobiles to be located.

The estimation error d_{err} of the above positioning technique has been determined by the authors from real-life measurements; d_{err} was the distance between the position estimated by the cell ID positioning method and the position obtained from a GPS measurement, which served as a reference and was assumed to be error free. Figure 8.5 shows the cumulative distribution function (CDF) of d_{err}, which we refer to as the estimation accuracy. According to Figure 8.5, 67% of the estimated position coordinates are less than 447 m away from the real position, and 95% are less than 1000 m away. The median corresponds to an estimation accuracy of 356 m.

Since the cell ID is also known on the mobile network side, this method can also be used in network-based positioning methods. The MSs know the cell IDs from evaluating the broadcast control channels (BCCHs) of the BTSs. Every MS tracks the BCCHs of up to seven BTSs in its neighborhood. This is usually done

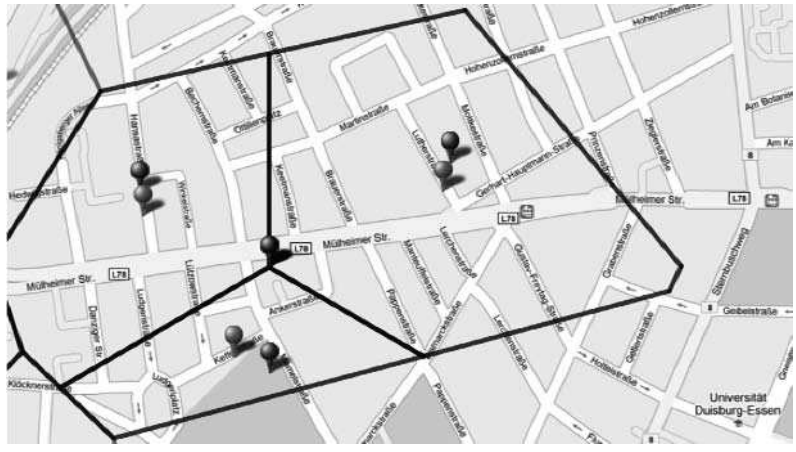


Figure 8.4 Cell geometry of a GSM network in Duisburg, Germany (© 2009 Google – Map data © 2009 Tele Atlas).

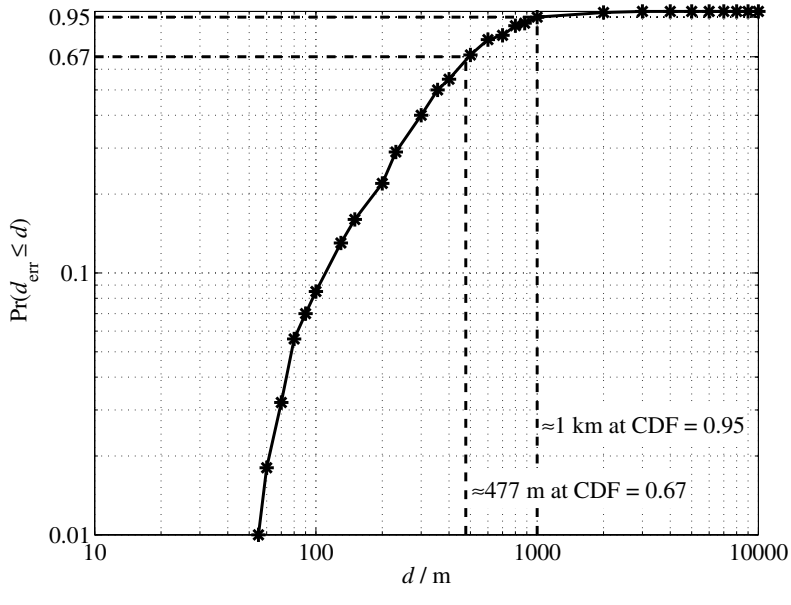


Figure 8.5 Localization accuracy measured in the area of Duisburg, using the cell ID positioning method.

Table 8.1 Mapping between the RSSI parameter in GSM and the actual value of received signal strength in dBm

RSSI	Received signal power p/dBm
0	$p < -110$
1	$-110 \leq p < -109$
⋮	⋮
62	$-49 \leq p < -48$
63	$-48 \leq p$

to allow appropriate preparations for handover. Besides the cell IDs, the tracking and measurement of the BCCHs provides values of the received signal strength indicator (RSSI).

8.3.2.2 RSSI

The RSSI is a six-bit value that indicates the short-time-averaged power of the BCCH received by the MS. Since an MS tracks the BCCH from up to seven BTSs in its neighborhood, it also determines up to seven RSSIs. The RSSI has a resolution of 1 dB. For instance, an RSSI value of 63 means that the BTS's radio signal is received with a received power of -48 dBm or more. An RSSI value of 62 indicates a signal power between -48 dBm and -49 dBm. The lowest RSSI value is 0, which corresponds to a received power of -110 dBm or below. Table 8.1 shows the mapping between RSSI and received signal power (3GPP 2005).

If the ERP of a BTS antenna is known and the MS measures the RSSI of the radio signal transmitted from that BTS, then the MS can calculate the attenuation of the mobile channel from BTS to MS (see Section 4.3.2). The channel attenuation is a function of the distance and can therefore be used to calculate an estimate of the distance between the BTS and the MS. There are several radio propagation models and path loss models, describing the relation between the channel attenuation and the distance between the radio transmitter and receiver. Common models which can be applied to cellular networks are the Okumura Hata model, the COST Hata model and the COST Walfish Ikegami model (see Section 3.2.1 or, for more detail, CCIR (1989)).

Owing to the channel impairments previously discussed in Section 3.2.1 and the consequent statistical properties of the noise, RSS measurements are unreliable for performing accurate positioning in GSM. Generally, the mobile-channel attenuation is assumed to be log-normal distributed (CCIR 1989). With ERP in dB and RSS in dBm, the mean logarithmic attenuation a in dB can be estimated from

$$a(d) = \text{ERP} - \text{RSSI}(d) - 110.5 \text{ dB}, \tag{8.20}$$

d being the distance between the mobile station and the BTS.

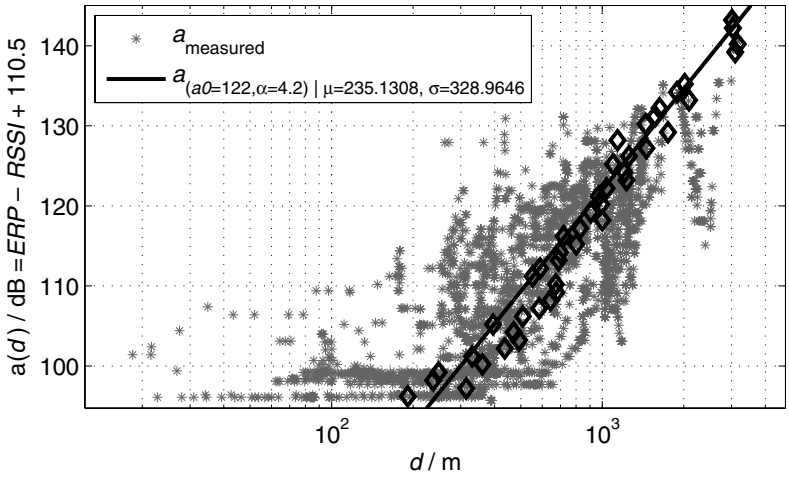


Figure 8.6 Channel attenuation obtained from RSSI measurements in the area of Duisburg.

Figure 8.6 shows the estimated attenuation, calculated after RSSI measurements, as a function of the distance d in the area of Duisburg. The solid line shows the logarithmic approximation to the relation between distance and attenuation which minimizes the absolute error of the estimated distance d. The mean error of this approximation is about 238 m. This estimate can be used to improve the localization when several RSSIs from different BTSs are known.

8.3.2.3 Mobile-assisted TOA

Since GSM uses TDMA, the radio signals from the MSs must reach the BTS in certain time slots. To allow accurate synchronization, the MSs must know the signal propagation delay of the mobile channel from MS to BTS. The timing advance (TA) is a six-bit value which indicates the signal propagation delay from the MS to the BTS and back – the so-called round-trip time. The latter is quantized in bit periods, i.e., the signal propagation time from BTS to MS and back is TA bit periods. In GSM, the bit period is (3GPP 2009)

$$t_{\mathrm{b}} = \frac{48}{13}\,\mu\mathrm{s} \approx 3.69\,\mu\mathrm{s}. \tag{8.21}$$

If we assume free-space propagation or LOS between the MS and the BTS, the distance d_{TA} between the MS and the BTS can be estimated as

$$d_{\mathrm{TA}} = \frac{t_{\mathrm{b}} c}{2} \cdot \mathrm{TA} \approx 554\ \mathrm{m} \cdot \mathrm{TA}, \tag{8.22}$$

where c is the speed of light. However, in most cases NLOS channels must be considered (see Chapter 7). This is usually the case in urban areas, where the radio

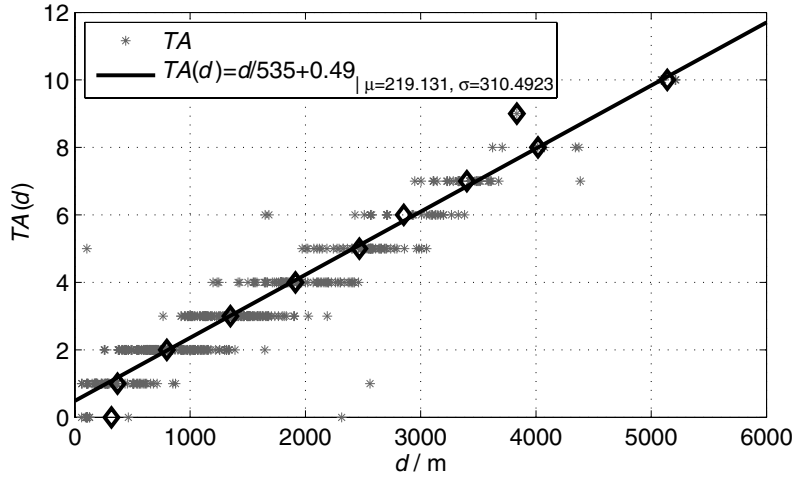


Figure 8.7 Measured timing advance and measured distance in the city of Duisburg.

signals often reach the receivers after reflections and scattering. This situation leads to an increased signal propagation delay and, consequently, to an overestimated distance d.

Along with cell IDs, the authors also measured the TA with a mobile phone in the area of Duisburg. The coordinates of the measurement points were determined by using a GPS-enabled device. The distances between the measurement points and the BTSs were calculated and served as reference distances to calculate the TA-based distance estimation error. Figure 8.7 illustrates the relation between the measured TA and the distance d. The solid line shows the linear approximation which minimizes the absolute estimation error. In field measurements, the average of the absolute distance estimation error was about 217 m when the linear approximation in Figure 8.7 was used.

The TA can generally be used to increase the localization accuracy. However, since it can be measured only when a dedicated channel is allocated, for example when a call has been initiated, the TA method is only conditionally applicable to positioning applications. On the other hand, the cell IDs and the RSSIs can always be queried from the mobile, and allow localization with increased accuracy compared with network-based methods without mobile assistance. The accuracy can be further improved by averaging consecutive measurements. Figure 8.8 shows the cumulative distribution function of the localization error d_{err} for mobile-assisted localization. Similarly to Figure 8.5, Figure 8.8 shows the CDF of d_{err}. According to Figure 8.8, 67% of the estimated position coordinates are less than 150 m away from the real position, and 95% are less than 300 m away. The median corresponds to an estimation accuracy of 117 m.

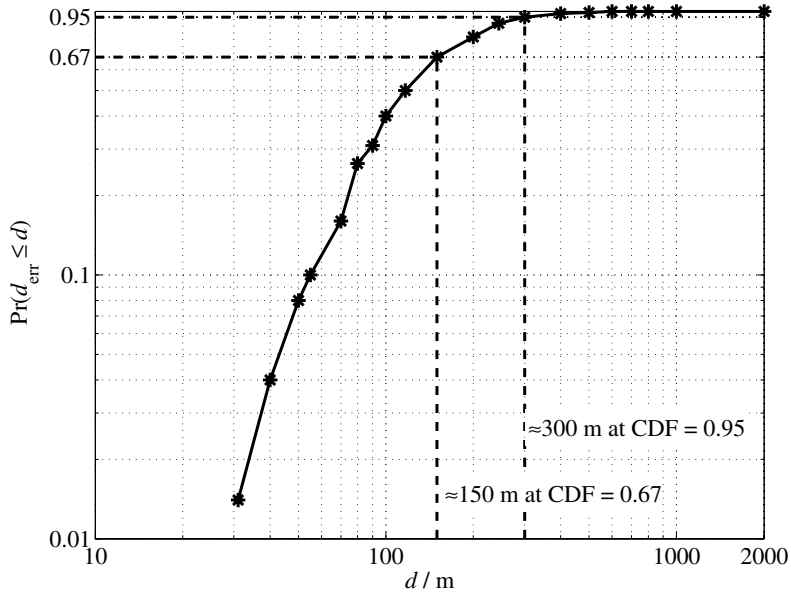


Figure 8.8 Localization accuracy measured in the area of Duisburg, using the mobile-assisted TOA positioning.

8.3.2.4 Accuracy and reliability

Selected results obtained by the authors are shown in Figures 8.9, 8.10 and 8.11. Figure 8.9 refers to measurements in Duisburg, and Figures 8.10 and 8.11 present a comparison of the positioning accuracy between network-based GPS-free positioning and mobile-assisted GPS-free positioning.

Figure 8.9 shows the map of the measurement area in Duisburg. The solid line shows the route taken and the positions estimated using GPS positioning. The 23 markers refer to "virtual center" positioning. The speed of the mobile was around 45 km/h and no temporal averaging of the measured GSM data was used. The quality of the measurements was improved in a second step by taking account of low mobility, for example at typical pedestrian speeds, and by taking the measurements of the candidate BTSs into account, for example by using linear estimation techniques. Let X_i be the position of the virtual center associated with the ith BTS and let $RSSI_i^{\mathrm{BTS}}$ be the received power value associated with the ith BTS and measured at the mobile; then the position of the mobile can be determined from

$$X = \frac{1}{\sum_{\forall i} \mathrm{RSSI}_i^{\mathrm{BTS}}} \cdot \sum_{\forall i} \mathrm{RSSI}_i^{\mathrm{BTS}} \cdot X_i . \tag{8.23}$$

This scheme is referred to as the "virtual center weighted" scheme. A further improvement can often be obtained by temporal averaging of consecutive positioning

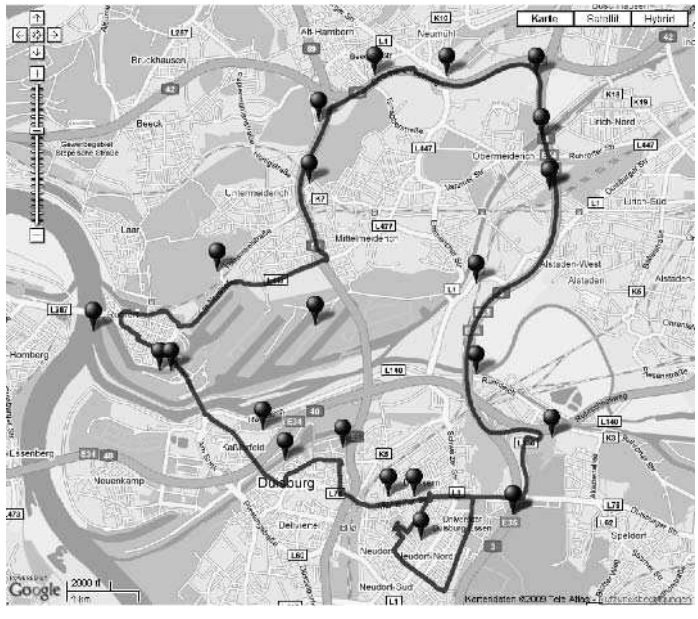


Figure 8.9 Measurement route in Duisburg (© 2009 Google – Map data © 2009 Tele Atlas).

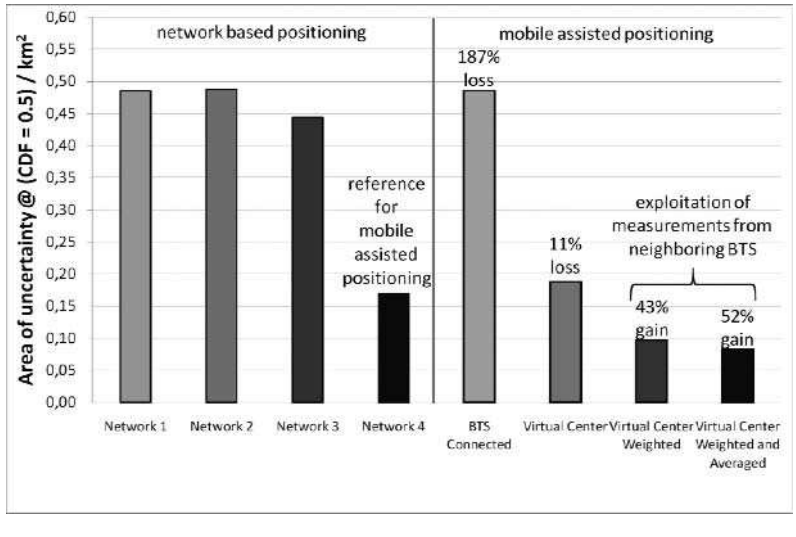


Figure 8.10 Median area of uncertainty (urban area).

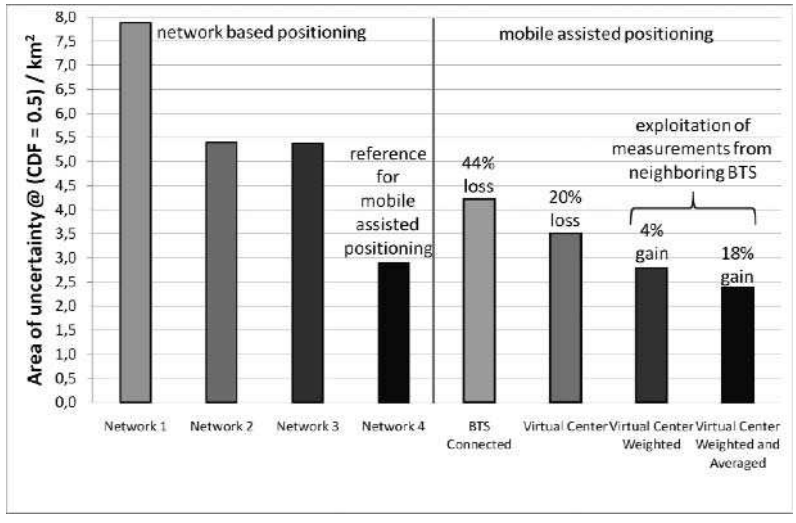


Figure 8.11 Median area of uncertainty (rural area).

results. The latter scheme is termed the "virtual center weighted and averaged" scheme (Waadt et al. 2008).

Figures 8.10 and 8.11 show results for the positioning of mobiles in GSM networks in urban and rural areas. Only cell ID and RSSI measurements were used. The measurements were taken in Duisburg and in rural areas between Duisburg and the Dutch border. The results are depicted as the "area of uncertainty," i.e., the area of a circle around the estimated location of a mobile with a radius representing half of the maximum positioning error. The positioning error is equal to the difference between the GPS coordinates of the mobile and the estimated coordinates of the mobile. Mobile-network-based positioning by four network operators was considered (Network 1 to Network 4). It was found that mobile-assisted positioning was often superior to mobile-network-based positioning and increased the localization accuracy by 52% in urban areas and 18% in rural areas.

Simple mobile-assisted GPS-free positioning is possible at a reasonably low implementation complexity and viable at the same time, when we consider the results in terms of positioning errors. Such mobile-assisted GPS-free positioning might be considered as an interesting alternative to mobile-network-based GPS-free positioning.

8.3.3 UMTS

8.3.3.1 3GPP standardization

Since LBSs are increasingly in demand, 3GPP has produced standards documents, which specify the location services (LCS) featured in GSM and in UMTS and the system evolved from it, the evolved packet system (EPS). Most of the mechanisms

to support mobile LCS for operators, subscribers and third-party service providers are presented in 3GPP (2008b,c,d). The standardized LCS may be considered as consisting of service capabilities that enable the provision of location applications.

3GPP has identified four categories of usage of location services (3GPP 2008b):

- The *Commercial LCS* will typically be associated with applications that provide value-added services to subscribers. This may be, for example, a directory of restaurants in the surroundings of the MS, together with directions for reaching them from the current location.

- The *Internal LCS* will typically make use of the location information of the user equipment for accessing internal operations of the network. These may include, for example, location-assisted handover, traffic and coverage measurements (see Section 2.4).

- The *Emergency LCS* will typically allow subscribers who place emergency calls to be located. This service may become mandatory in some jurisdictions. In the United States, for example, this service is already mandatory for all mobile voice subscribers.

- The *Lawful Intercept LCS* will use location information to support legally required or sanctioned services, for example localization for law enforcement agencies, as required by the local jurisdiction.

As discussed in the earlier sections of this chapter, the capability to determine the geographic location of an MS is provided by making use of the radio signals. The location information may be requested by and reported to a client application associated with the MS, or by a client within or attached to the core network. Furthermore, the location information request may ask for the speed of the MS as part of the positioning information. The latter is reported in accordance with the specifications in 3GPP (2008d).

The LCS feature will be based on the following positioning methods (3GPP 2008a): (1) cell ID; (2) Observed Time Difference of Arrival–Idle Period Downlink (OTDOA-IPDL); (3) uplink–time difference of arrival (U-TDOA); and (4) network-assisted GNSS methods, also termed A-GNSS-based positioning. Since the cell-coverage-based positioning method has already been presented in Section 8.3.2, only the OTDOA-IPDL, the U-TDOA and the A-GNSS-based positioning methods will be presented in the following subsections.

8.3.3.2 OTDOA-IPDL

The OTDOA-IPDL positioning method (3GPP 2008a) involves measurements made between the MS and the location measurement unit (LMU) of the UMTS Terrestrial Radio Access Network (UTRAN). The simplest case of OTDOA-IPDL positioning is the one without idle periods. In this case, the method can be referred to as simply

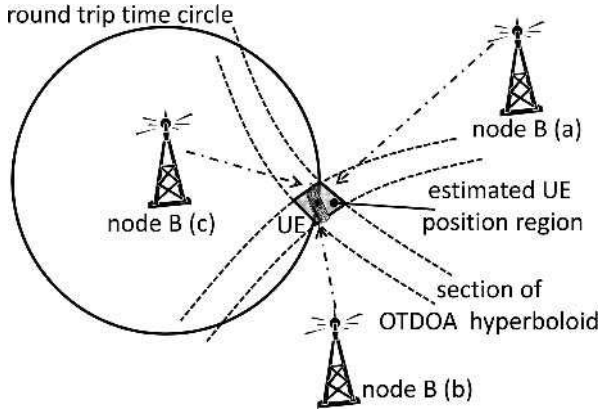


Figure 8.12 OTDOA positioning (UE = user equipment).

OTDOA positioning. A node B (e.g., a BS) may provide idle periods in the downlink, in order to potentially improve the hearability of neighboring nodes B.

The primary standard OTDOA measurement is the observed time difference between two system frame numbers (SFNs), which is observed by the MS. These measurements, together with other information concerning the surveyed geographic position of the transmitters and the relative time difference (RTD) of the actual transmissions of the downlink signals, may be used to calculate an estimate of the position and, optionally, the speed of the MS. Each OTDOA measurement for a pair of downlink transmissions describes a line of constant time-of-arrival difference, yielding a hyperboloid in three dimensions. The MS's position is determined by the intersection of the resulting lines for at least two pairs of nodes B. The accuracy of the position estimates made with this technique depends on the precision of the timing measurements and the relative position of the nodes B involved, and it is also subject to the effects of multipath radio propagation. The best results are obtained when the nodes B equally surround the MS. This is illustrated in Figure 8.12 (3GPP 2008a).

8.3.3.3 U-TDOA

The U-TDOA positioning method (3GPP 2008a) is based on network measurements of the TOA of a known signal sent from the MS and received by at least four LMUs. The signal propagation time from an MS to an LMU is proportional to the distance between the MS and the LMU. The difference between the times of arrival at two LMUs defines a hyperbola. The MS's location can be estimated from the intersections of these hyperbolas. An advantage of this method is that it does not require knowledge of the time that the MS transmits, nor does it require any new functionality in the MS.

Figure 8.13 illustrates how the time delay is determined (3GPP 2008a). Two receivers, RX1 and RX2, represent the fixed locations of two U-TDOA-capable LMUs in a network. The pairs of intersecting circles centered around RX1 and RX2

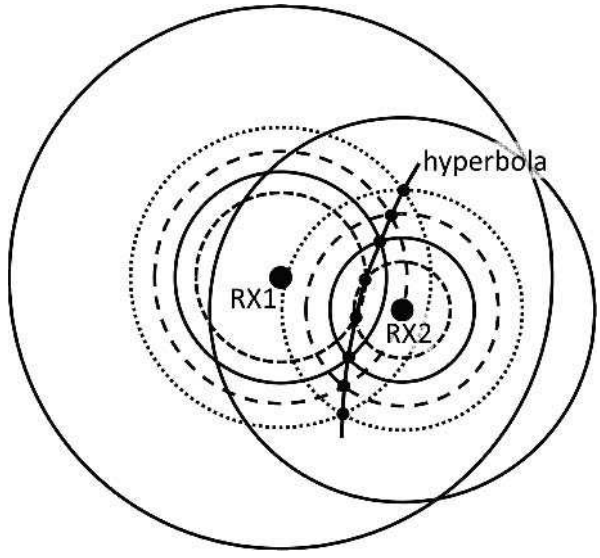


Figure 8.13 Deriving the hyperbolic function for the time delay.

have radii that represent the times t^{tr} it takes a signal to travel from an MS transmitter to each U-TDOA-capable LMU. The difference between the radii, however, is a constant equal to the difference between t_1^{tr} for the first U-TDOA-capable LMU and t_2^{tr} for the second LMU. This difference, $(t_1^{tr} - t_2^{tr})$, is the value which the U-TDOA system measures. When the points of intersection of the pairs of circles are plotted and connected, the resulting shape is a hyperbola. The latter, defined by a reference site in comparison with another site, constitutes a baseline. If one constructs a similar hyperbola for another pair of U-TDOA-capable LMUs receiving transmissions from the same MS, that hyperbola will intersect with the first hyperbola at two points, yielding two possible locations for the MS. A third hyperbola will yield a unique location for the MS, so that a minimum of four reception sites is needed to obtain a unique location estimate.

8.3.3.4 A-GNSS-based positioning

The A-GNSS-based positioning method makes use of MSs which are equipped with radio receivers capable of receiving GNSS signals (3GPP 2008a). Examples of GNSS include GPS, GALILEO, GLONASS, SBASs and the Quasi-Zenith Satellite System (QZSS). In this context, different GNSSs can be used separately or in combination to perform the location of an MS. Assisted GPS is discussed in more detail in Section 8.6.3.

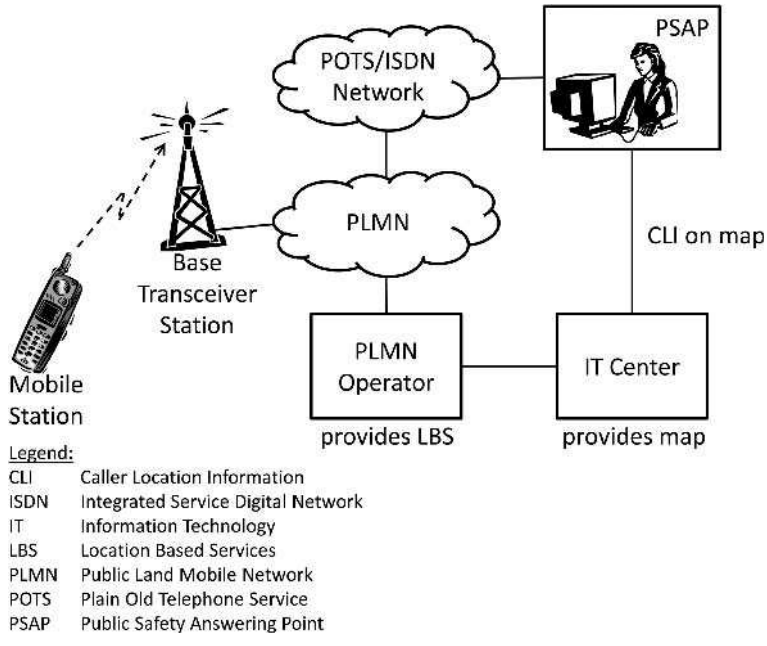


Figure 8.14 Emergency caller localization system in a mobile network.

8.3.4 Emergency applications in cellular networks

A classical application of cellular positioning is the localization of the origin of an emergency call. The purpose of this application is to allow the mobile network to route the emergency call to the closest public safety answering point (PSAP) and to immediately inform the PSAP call-taker about the location of the emergency caller and where exactly to send emergency services. The facility that a network offers to localize the origin of an emergency call is often referred to as caller location. In North America, this service is also called Enhanced 9-1-1 (or E911).

In wired telephone networks, caller location has been employed since the late 1960s, with the result that there is only a small delay between the reception of an emergency call at the PSAP and the arrival of emergency services at the location of the emergency caller. Owing to the lack of caller location services for cellular networks, this delay has increased significantly with the spread of mobile phones since the 1990s. Mobile networks and PSAPs provide caller location service in a fully automated fashion, i.e., the location of the emergency caller is automatically displayed to the PSAP call-taker. However, this service has not been implemented in all networks and PSAPs. The eCall project in the eSafety initiative of the European Commission is intended to provide a unique, fully automatic caller location service for all European streets (European Commission 2005).

Figure 8.14 shows a typical system framework for a caller location service. The emergency call, originating at an MS, is recognized by the operator of a public

land mobile network (PLMN). Today, PLMN operators typically use network-based cellular positioning techniques to determine the location of an emergency call. Then, the network operator routes the call through the PLMN and the public switched telephone network (PSTN) to the closest PSAP and provides the caller location information (CLI). The evaluation of the location information and the presentation of the mobile subscriber's location on the map are often accomplished by a third-party service provider, illustrated as the information technology (IT) center in Figure 8.14.

8.3.5 Drawbacks when applied to mobile positioning

Mobile positioning in cellular systems has many advantages. These include a short time to the first position fix, good coverage and low cost owing to the fact that the information to be exploited is gathered in the mobile network itself. However, there are also some disadvantages. A major drawback is the limited accuracy. It has been shown that the accuracy is typically not better than approximately 100 m at a reasonable complexity. Furthermore, the accuracy varies greatly with the density of nodes B: in the case of urban areas with a dense network, the accuracy is reasonably good, whereas in rural areas, with a much lower number of nodes B, the absolute accuracy decreases. The results obtained in the previous sections show that the relative accuracy can achieve values of the order of approximately 15–20% of the cell radius.

8.4 Wireless Local/Personal Area Network Positioning

8.4.1 Solutions on top of wireless local networks

One reason for the limited accuracy and reliability of localization in satellite and cellular positioning systems lies in the large distances between the communication nodes, i.e., the transmitters and the receivers. Owing to these large distances, small relative measurement errors may lead to large absolute position estimation errors. Those relative measurement errors include impacts from NLOS propagation of the radio signal in cellular networks, and erroneous estimations of the signals' ionospheric delays in the case of satellite positioning. Positioning in WLANs, such as Wi-Fi and UWB systems, therefore seems to be a promising approach, because the distances between communication nodes are small compared with cellular or satellite-based positioning systems. On the other hand, the applicability of positioning in WLANs is limited owing to the small dimensions of the network. There are, however, numerous applications of positioning in WLANs in all branches of the industry, including home entertainment, consumer electronics, the automotive industry, public transportation and heterogeneous cellular networks. Hence, it is not surprising that industry and academia are jointly developing short-range wireless communications systems with integrated localization and tracking capabilities. For instance, the European Ultra-Wideband (EUWB) project, supported by the European Union's Seventh Framework Program for Research and Technological Development,

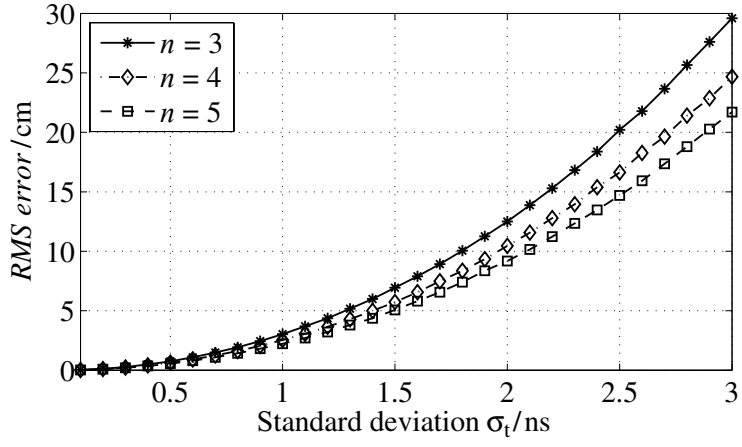


Figure 8.15 RMS localization error in UWB scenarios.

is investigating UWB systems with data rates ranging up to gigabits per second and precise real-time location and tracking (LT) (Zeisberg and Schreiber 2008). Accurate real-time positioning is feasible in UWB systems because of the unique features of ultrawide radio frequency band allocation (Zeisberg and Schreiber 2008).

8.4.1.1 UWB

UWB is a short-range communication system that uses a large frequency bandwidth at a low spectral power density. One goal of UWB was to transmit in a way which does not interfere with narrowband communication systems in the same frequency band. There are various standardization and industry associations, specifying UWB in the physical and MAC layers of communication systems. Examples of UWB applications include high-data-rate services over short distances, such as indoor video streaming and wireless personal area networks (WPANs) (Heidari 2008), and low-power, down-market devices that allow moderate data rates and positioning (IEEE 2005).

Interesting scenarios for the application of LT in the EUWB project include both small-scale environments and spacious indoor environments such as shopping malls, train stations, airports, exhibition centers and sports stadiums. Providing data transmission and LT capabilities simultaneously, a single UWB network can be used to interconnect several different sensors and to provide location information to mobile users. In the same manner as in automotive navigation systems, users can locate themselves on their mobile device and find their way to a shop, a check-in desk or a seat on an aircraft. Furthermore, users could also get location-related information through the UWB network, such as points of interest. Also, the network can evaluate knowledge about users' locations; when this is sent to a cellular network, it can improve RRM performance (see Section 2.4). In small-scale environments, indoor

or in-cabin, LT can be used to localize a wireless key, for instance, or to identify whether a particular seat on an aircraft has been occupied or not.

Cross-layer functionality is required to accomplish LT applications. An adequate communications protocol must comprise three steps: (1) synchronization of the nodes, (2) localization of a mobile device and (3) triangulation and mapping.

Accurate positioning in wireless local area networks is based on time-of-flight or TOA measurements. Similarly to cellular and satellite positioning, these measurements first require synchronization of the nodes, including the infrastructure, i.e., the access points (APs), and the MS. After synchronization, the measurement and calculation of the distances between the access points and the mobile station can be accomplished. Finally, the location of the mobile station can be computed by an access point and conveyed to the other nodes, if required.

The accuracy of the aforementioned LT procedure has been simulated for a scenario with $n = 3$, $n = 4$ or $n = 5$ access points, equally spaced on a circle with a radius of 1 m. The mobile station was located in the middle of the configuration. Figure 8.15 shows the root mean square (RMS) error of the position estimation; it is in the range of centimeters and depends on the standard deviation σ_t of the TOA measurements.

8.4.1.2 Bluetooth

Bluetooth is a short-range technology in which two different approaches are used for positioning. One approach assumes that positioning or tracking is to be performed on a large scale, i.e., a large area covered by several access points, where the granularity of the positioning system is given by the positions of the access points (Gonzalez-Castano and Garcia-Reinoso 2002; Pels et al. 2005). Another approach aims at estimating position on a subcell-scale level, i.e., the granularity is finer than the access point coverage (Bandara et al. 2004; Feldmann et al. 2003; Figueiras and Schwefel 2008; Figueiras et al. 2005).

On large scales, the typical information used to localize an MS is cell ID data. The use of this cell ID technique limits such systems to a granularity given by the coverage of the APs. Consequently, the magnitude of the achieved accuracy is given by the cell range in the Bluetooth communications, which depends on the class of the device: class 1 devices have a typical range of 100 m, whereas the class 2 devices have a typical range of tens of meters.

For subcell localization, the RSSI information defined by the Bluetooth specifications is the key to estimating positioning information. The RSSI is an octet defined in the Bluetooth specifications and ranges from -128 to 127 in single integers. The value 0 specifies the so-called golden receive power range, above which the RSSI values are positive and below which the RSSI values are negative. According to Bluetooth SIG (2003), the golden receive power range is defined as the range of values from the lower threshold, loosely defined as placed between -56 dBm and 6 dB above the receiver sensitivity, to the upper threshold, placed 20 dB $\pm$ 6 dB above the lower threshold. Figure 8.16 shows these levels.

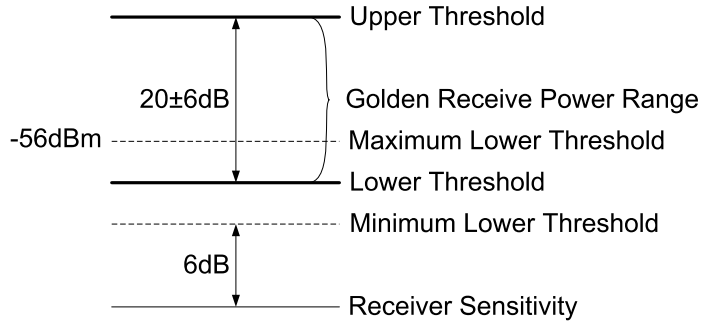


Figure 8.16 RSSI thresholds in Bluetooth.

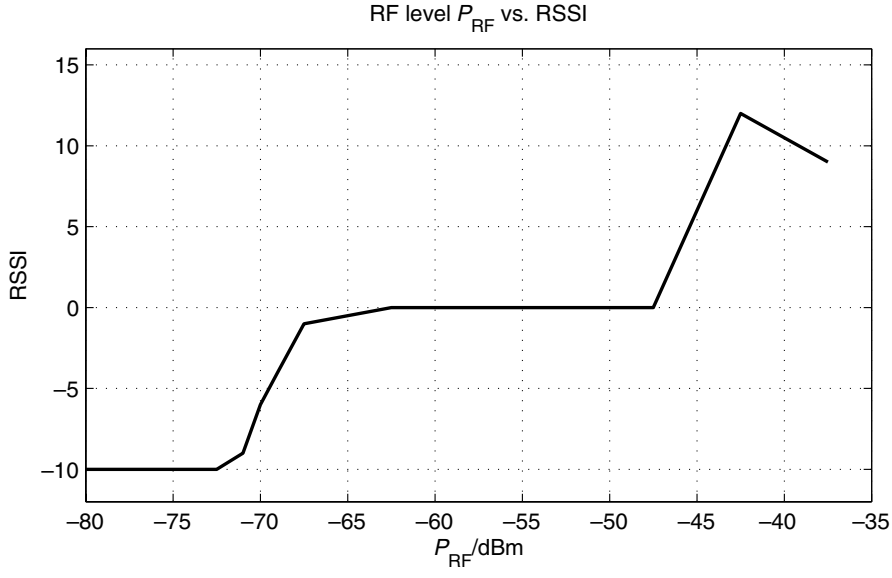


Figure 8.17 RF level vs. RSSI.

This means that the golden receive power range has a dynamic range of 20 dB ± 6 dB, within which the RSSI value is 0. Since the threshold levels are left open for the manufacturers of the Bluetooth chips, the translation of RSSI values into actual received signal strength values depends on the particular Bluetooth chip. Nonzero values of RSSI define the magnitude in dB by which the received signal strength is below or above the lower or upper threshold, respectively. Figure 8.17 shows an example of a possible relation between the RSSI and the received signal strength.

Referring to the Bluetooth specifications, the RSSI value is tightly connected to the power control mechanism that adapts the link conditions in a Bluetooth communication. Careful analysis reveals that when power control is active, the

transmission power used by the transmitter is not necessarily known at the receiver, meaning that the estimation of path loss may be complex to perform. A common solution is to run the positioning mechanisms when the Bluetooth device is in the inquiry state.[1] The reason is that during the inquiry state, the power control mechanism is inactive and the transmission is done at the maximum transmission power specified by the Bluetooth standards.

8.4.1.3 WLAN (Wi-Fi)

There is already a commercially available positioning system for WLAN called Loki (Skyhook 2009), which works similarly to the cell ID method. Instead of the cell IDs of BTSs, it uses the MAC addresses of nearby WLAN access points to estimate a position. Since the WLAN cell size is much smaller than the cell size in a cellular mobile network and a WLAN client can observe several MAC addresses, the WLAN positioning service outperforms the cell ID method with respect to its accuracy. Loki requires access to a database of registered WLAN access points and their locations. Therefore, it is a requirement that every user of this service registers their own WLAN access point with this publicly available database.

8.4.2 Dedicated solutions

The positioning methods discussed above make use of communication systems whose main purpose is data transmission rather than positioning. Dedicated positioning solutions, in contrast, are expressly designed for positioning applications and require specific hardware for this purpose. The only signals exchanged by the transmitter and receiver are the positioning signals.

8.4.2.1 RFID

An old but newly reemerged technology is RFID, using so-called RFID tags for identification and tracking. RFID tags can also be used to increase the accuracy and reliability of popular positioning systems such as the GPS. For instance, we could have a number of fixed installations on a roadway. Every installation includes an RFID tag along with information about its accurate position. The mobile receiver contains a RFID reader to read this location information in order to provide more accurate positioning, especially in tunnels and in downtown areas, where satellites are often obstructed.

The system described in Chon et al. (2004) was proposed as a complementary outdoor system to GNSSs, and can be summarized as follows. RFID tags need to be installed on a road in a way that maximizes the coverage and the accuracy of positioning (see Figure 8.18). During installation, necessary information, such as the coordinates of the location where the tag is installed, needs to be written into each

[1] The inquiry state is the initial state that a Bluetooth device enters in order to discover other Bluetooth devices with the intention of establishing a communication link.

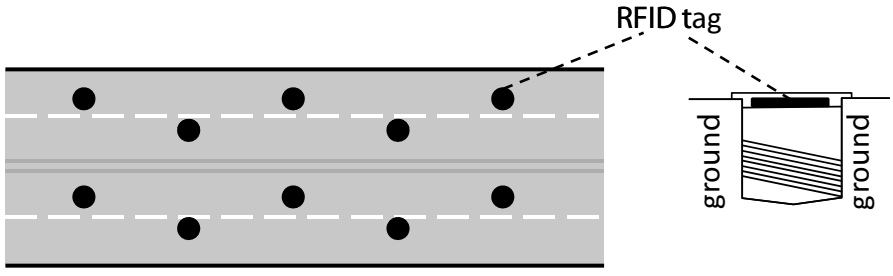


Figure 8.18 Road-mounted RFID tags.

tag. The accuracy of this position information is very critical for this technique to be successful. The position information can be acquired by using, for example, the DGPS. Vehicles then need to be equipped with an RFID reader that can communicate with the tags on a road. No matter how accurate the RFID positioning is, the system provides the positions where the tags are (Chon et al. 2004). Therefore the vehicles need also to be equipped with a GPS receiver and inertial sensors such as a gyroscope for positioning when there are no tags around. While being driven, the vehicles constantly monitor for the presence of a tag. On detection, the reader retrieves the information from the tag, including its coordinates. Figure 8.18 shows RFID tags deployed on a road. The circles on each lane represent a tag and the tag itself is enclosed in a special-purpose reflector, a so-called "cat's eye."

8.4.2.2 Infrared

In 1977 the Institute of Electrical and Electronics Engineers (IEEE) redefined radar as "an electromagnetic means for target location and tracking," which includes electro-optical devices such as laser radars (lidars) and laser rangefinders in general. Relative to microwave and millimeter-wave systems, electro-optical sensors are characterized by extremely short wavelengths, thus affording much higher resolution but suffering greater attenuation due to atmospheric conditions. The optical region of the electromagnetic spectrum is broken up into the ultraviolet, visible and infrared domains. This subsection concentrates on positioning in the infrared domain.

A typical model of an electro-optical ranging system comprises five main components: (1) the optical source, typically a light emitting diode or a laser diode; (2) a mechanism to modulate the optical output from the source with the signal to be transmitted; (3) the transmission medium; (4) a photodetector, which converts the received optical power back into an electrical waveform; and (5) electronic amplification and signal processing to recover the signal and to present it in a suitable form.

Infrared ranging is performed by using either continuous-wave signals, narrowband signals (Eltaher 2009) or pulse-train-based transmission, which uses, for example, short impulses or impulse compression concepts (Zimmermann 1991).

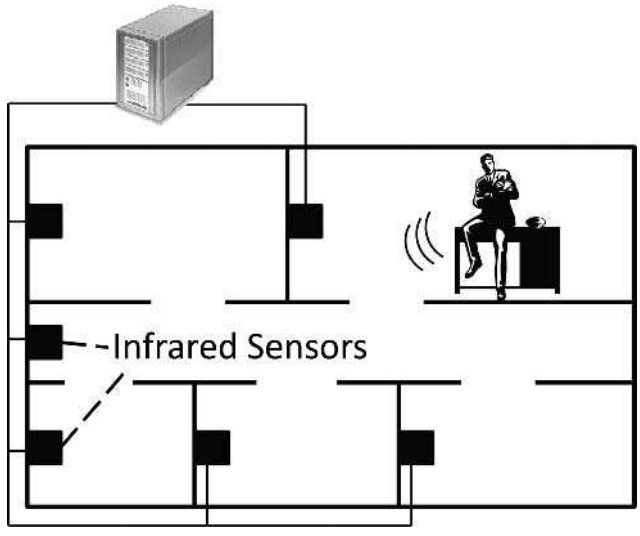


Figure 8.19 Infrared positioning.

Typically, the range of such systems is limited to a few tens of meters; however, the accuracy can be as low as a few tens of micrometers. For a survey of infrared technology, the reader is referred to Eltaher (2009).

Figure 8.19 shows an infrared positioning scenario in the case of a small indoor environment. Each room is equipped with an infrared sensor. These infrared sensors are connected to a server that carries out the location estimation. A mobile infrared device, for example worn by a human, transmits infrared signals, which are detected by at least one particular sensor in the vicinity of the mobile infrared device. The detected signal is then used to evaluate the position of the mobile infrared device.

8.4.2.3 Ultrasound

For indoor applications, ultrasonic positioning systems have emerged which make use of a combination of RF and ultrasound technologies. A number of fixed nodes, installed on walls and ceilings, transmit their location information via RF signals and at the same time ultrasonic pulses, which are synchronized with the RF signals. A mobile node correlates the received RF signal and the ultrasonic pulses. By taking the difference between the propagation speeds of RF (the speed of light) and ultrasound (the speed of sound) into account, it can estimate the distances to different fixed nodes and then its own position. Examples of commercial solutions have been presented by Intersense (Wormell et al. 2007) and Olivetti (Ward et al. 1997).

A specialty of the work on ultrasonic positioning has been the estimation of the orientation of objects (Ward et al. 1997). If three mobile stations are placed on a rigid body at known noncollinear points, then, by calculating the positions of the devices, the orientation of the object can be found. This technique is most suitable

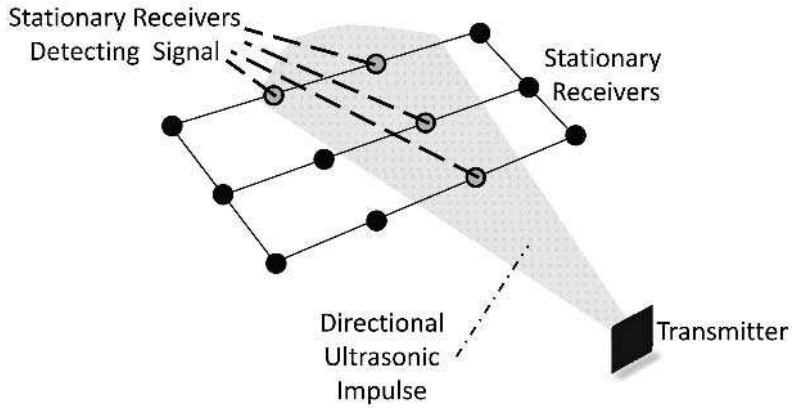


Figure 8.20 Ultrasonic positioning and orientation estimation.

for stationary or slow-moving objects, because the three devices will be located at different times – any intervening movement of the object will introduce errors into the calculated orientation. Alternatively, an object's orientation can be obtained using knowledge of the set of receivers that have detected ultrasonic signals from a mobile station (Ward et al. 1997). The directional transmission pattern of the ultrasonic pulse is known to be hemispherical. With this knowledge in combination with the known position of the mobile station, the set of sensors which receive the mobile station's ultrasonic pulse can be used to estimate the mobile station's orientation (Figure 8.20). If the device is rigidly affixed to an object at a known point and in a known orientation, an approximation to the object's orientation can then be deduced. This approach is most suitable when the tagging of the object with multiple transmitters could be cumbersome.

8.5 Ad hoc Positioning

In large isotropic networks, for instance, where the routing effort does not depend on the routing direction, the availability of position information enables routing without the use of large routing tables. This application is often referred to as the primary positioning application in ad hoc networks, and the GPS is often used (Ko and Vaidya 1998). In Section 8.3, we discussed positioning methods that are also applicable to ad hoc positioning. These rely on the nodes' capability to measure TOA, TDOA and RSS and on a knowledge of the base stations' locations. Decentralized, self-organized, scalable ad hoc networks do not contain fixed base stations. Locations must therefore be determined in a relative coordinate system. Algorithms for relative localization can be found in Capkun et al. (2002). These are based on the detection of the nodes in the ad hoc network and on the TOA method to estimate the distances between neighboring nodes. In a decentralized approach, every node measures the distances

to its neighboring nodes. The collected information about the network topology and the distances between the nodes is then communicated to all nodes in the network. Based on this knowledge, the relative positions can be estimated.

8.6 Hybrid Positioning

8.6.1 Heterogeneous positioning

The coexistence of different wireless access technologies creates a highly dynamic environment, which can be viewed as a heterogeneous network. To establish appropriate interoperability, heterogeneous positioning systems can be created to facilitate increased positioning accuracy and better coverage for LBSs. In this context, two different kinds of heterogeneous positioning can be distinguished.

The aim of the first kind of heterogeneous positioning systems is to extend conventional indoor positioning systems by introducing a homogeneous interface between them and outdoor positioning systems such as the GPS. In addition, when cellular networks, which have wide area coverage, high mobility, small bandwidth and, usually, moderate data rates, are combined with short-range systems such as WLANs, which typically have low coverage, low mobility, high bandwidth and, usually, high data rates, the positioning accuracy can be improved. This aspect is treated further in Section 8.6.2.

The aim of the second kind of heterogeneous positioning systems is to extend conventional outdoor positioning systems and to reduce the position acquisition times. For instance, this can be achieved by combining cellular positioning with GNSSs. In this case, the positioning coverage of GNSSs can be improved in cases where the LOS to the satellites is obstructed, for example. This aspect is treated further in Section 8.6.3.

8.6.2 Cellular and WLAN

Recently, the combination of cellular and WLAN radio interfaces in handheld devices has been introduced. Essentially, the position estimation techniques deployed in WLANs are the same as in the case of cellular systems and can therefore be reused in combined cellular and WLAN positioning.

Seamless LBS provisioning for users moving between outdoor and indoor environments requires a complex interaction between the different positioning systems. Zuendt et al. (2005) provided a concept that facilitates an open homogeneous location architecture which delivers the best possible positioning information to a mobile station. A conceptual architecture of a combined GSM and WLAN positioning scheme is depicted in Figure 8.21. This figure illustrates the case of a multistandard-capable MS which supports both GSM and WLAN radio interfaces. Positioning is carried out by enhancing the LBS data obtained from GSM by using the WLAN connectivity of the MS. In practice, the MS requests improved positioning via the WLAN access point, which conveys this request to the location

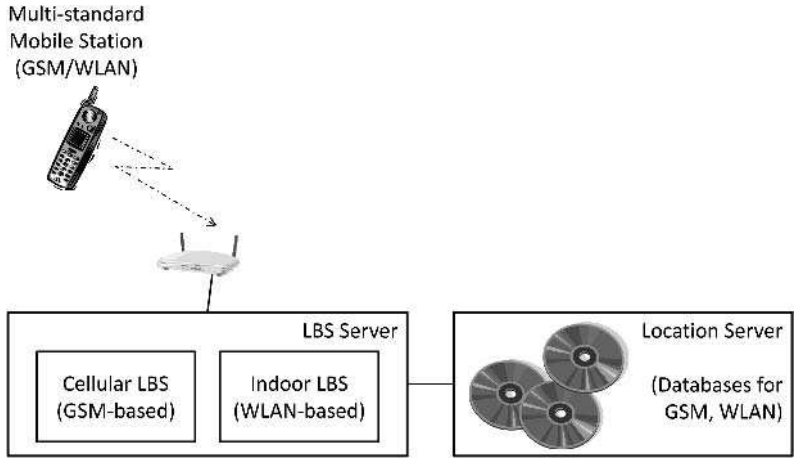


Figure 8.21 Conceptual architecture of combined cellular and WLAN positioning.

server. The latter provides both cellular and WLAN positioning features and is connected to the WLAN and GSM infrastructure databases. These databases hold information about the location of BTSs and WLAN access points. The location server computes the position of the MS by exploiting the data in the databases and conveys this piece of information back to the MS. In addition, the location server could also provide a link to a map showing the location of the MS. The accuracy of this scheme is limited mainly by the coverage of the WLAN access points. Zuendt et al. (2005) reported a typical estimation accuracy as low as 2 m. The drawback of this scheme lies in the fact that the establishment of a location database of access points is cumbersome and thus at present limits the deployment of WLAN positioning.

8.6.3 Assisted GPS

When a GPS system is designed to interoperate with cellular networks, the network assists the receiver to improve the performance with respect to the startup time, sensitivity and power consumption. In order to perform localization, the receiver must first know the orbital elements of the satellites. These data are transmitted with the GPS broadcast signal. The data rate of the GPS broadcast signal is only 50 bits per second. That is why the provision of a first position fix can take up to 12.5 minutes. Alternatively, the satellites' orbital elements can be transmitted via a cellular network connection, thus reducing the startup and acquisition times to a few seconds.

A-GPS is considered to be the first A-GNSS. It was initially proposed in the early 1980s. In A-GPS, the GPS receiver is provided with information that helps to reduce the time it takes the receiver to calculate its position. Also, the sensitivity can be improved by using this method when the receiver is unable to demodulate the GPS signals in low-SNR situations. Moreover, the rapid startup time allows the receiver

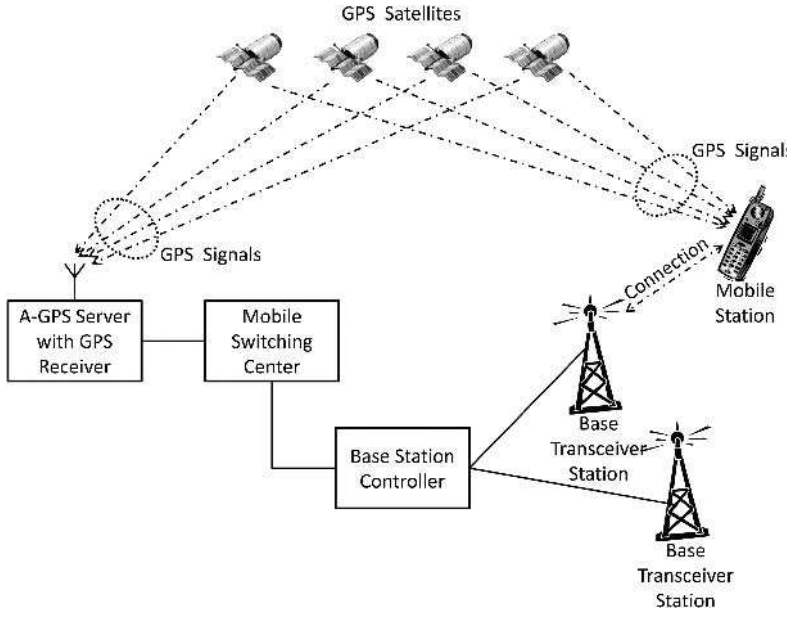


Figure 8.22 A-GPS conceptual architecture.

to stay in idle mode when positioning is temporally not needed. This reduces the receiver's power consumption.

The A-GPS concept is shown in Figure 8.22. The MS has a GPS receiver integrated into it. The A-GPS server has a reference GPS receiver, which is in LOS conditions and monitors the same satellites as are seen by the MS. Each A-GPS server provides information for various BTSs and has exact knowledge about the GPS signals that are available at the MSs. The A-GPS server is connected to the mobile switching center. The A-GPS server sends information to the MS by means of the cellular radio network. This includes information about the satellites which can be received in the corresponding area. The MS utilizes this information and as a result improves its positioning capability. A-GPS became truly relevant and needed only after the wireless E911 mandate was issued (FCC 1999).

8.7 Conclusions

GNSSs such as the GPS can provide precise positioning with global coverage, as long as the receivers have an LOS connection to at least four satellites. This constraint is the main disadvantage of the GPS, leading to limited reliability. Cellular positioning techniques have excellent reliability in urban areas. However, they are usually significantly less accurate than GNSSs. Nevertheless, they can potentially be improved by advanced mobile-assisted positioning. In particular, combination with

the upcoming wireless local-area-network positioning leads to improvement of the position estimation accuracy, and thus makes them competitive with GNSSs.

Acknowledgements

The authors gratefully acknowledge the support of Björn Steiger Stiftung Service GmbH and of Vodafone. Furthermore, the authors wish to thank their colleagues at the Lehrstuhl für KommunikationsTechnik, Universität Duisburg-Essen, for fruitful discussions, hints and helpful suggestions.

9

Cooperative Mobile Positioning

9.1 Introduction

As described in previous chapters, the underlying technologies used in mobile positioning systems are either *satellite-based* or *terrestrially based*, and differ from each other in terms of accuracy, coverage, cost, energy consumption, application environment (outdoor or indoor) and system impact. Satellite-based technologies are employed mainly for outdoor applications and are mainly represented by the GPS. Even though the latter is the most popular solution on the market because of its navigation applications, the introduction of mobile handsets with built-in receivers for the 3G system led to increased cost, size and battery consumption, and a long time for full market penetration (Sayed et al. 2005a). Furthermore, in severely handicapped environments, such as outdoor urban canyons and indoor environments, which actually represent the greatest interest of service providers, it is difficult, if not impossible, to obtain any sort of location information. This is due to the unfeasibility of having a clear view of at least four satellites, or to signal blocking and multipath effects. Current terrestrially based (GPS-free) technologies have the same drawbacks in multipath environments and in NLOS conditions, in which none of the costly accurate environmental information is available. Hence, investigations have started in connection with the fourth generation (4G) in order to define a solution that overcomes the drawbacks of the GPS and the current terrestrially based technologies and provides location information with a high level of accuracy *anywhere and anytime*. As described in Frattasi (2007), wireless cooperation gives designers the potential to achieve enhancements in terms of coverage, spectrum usage and energy consumption. In this chapter, we will discuss the impact of wireless cooperation on location estimation accuracy.

Cooperative localization is an approach that has been used in robot networks and wireless sensor networks (WSNs) in order to enhance the location estimation accuracy and thus enlarge the range of applications of such networks. In practice,

Mobile Positioning and Tracking João Figueiras and Simone Frattasi
© 2010 John Wiley & Sons, Ltd

in addition to using measurements from known to unknown devices, measurements between unknown devices are also exploited, which permits a more robust and precise localization to be achieved. In particular, WSNs employ clustering as a technique to reduce the volume of data to be transmitted to the remote station by aggregating location measurements from co-located sensors in a master device, which then takes care of the long-range transmissions. Recently, the cooperative localization approach has been ported to wireless mobile networks, where the actors are not anymore robots or sensors but users, with their mobility patterns and social behaviors. In this context, we have coined the expressions *cooperative mobile positioning* and *cooperative augmentation system* to refer to cooperative localization applied to wireless mobile networks and to localization systems employing such an approach, respectively.

Cooperative mobile positioning is becoming increasingly important as a promising new branch of wireless location, in which several research directions are being explored (e.g., positioning, tracking, data fusion and clustering). Its identifying concept is the exploitation of likely reliable mobile-to-mobile measurements to increase the location estimation accuracy of a satellite-based or terrestrially based system, which would have been provided otherwise only by fixed-to-mobile measurements. In this chapter, we use the framework of cooperative mobile positioning framework and introduce the COMET as an innovative solution for position determination in 4G wireless networks. COMET is supported by the "ad coop" architecture specified in Frattasi (2007). Specifically, peer-to-peer (P2P) communications are exploited in a mesh fashion within cellularly established clusters for cooperation-aided localization purposes. Upon receiving a location information request from an MS, the home BS forms a cluster in the surroundings of that MS, where the latter can be chosen as the cluster head (CH) and its neighbors as cluster members (CMs). While the available BSs perform time/time-difference-based and angle-based measurements for each BS–CM link, each CM performs range-based measurements for each CM–CM link. Finally, the positions of all the CMs are obtained by novel data fusion methods, which appropriately combine the available long- and short-range location information. The numerical results presented in Section 9.4 show that thanks to the cooperation established within a group of mobiles, our proposal has the potential to enhance the location estimation accuracy with respect to conventional terrestrial positioning systems in stand-alone cellular networks.

The rest of the chapter is organized as follows. Section 9.2 presents a brief survey of cooperative localization in robot networks and WSNs, and introduces the cooperative mobile positioning approach and its overall benefits. Section 9.3 illustrates the data fusion and filtering techniques of Chapter 5, which have been adapted to cope with the cooperative framework. Section 9.4 describes COMET, an example of a cooperative augmentation system (CAS) in a cellular network, with its system architecture and its data fusion methods; in particular, Section 9.4.3 discusses its performance, which has been evaluated via computer simulations. Finally, our concluding remarks are made in Section 9.5.

9.2 Cooperative Localization

9.2.1 Robot networks

Robotics is a field that has been around for decades and ranges from the simplest domestic appliances to the more sophisticated robots used in manufacturing plants and warehouses and the ones used in the realm of biomedical engineering, for example for teleguided microsurgery. In particular, when we move from fixed to mobile robots, an avenue of other applications pops up (MobileRobots 2009; Parker 1996): delivery, guidance, mapping, patrolling, environmental monitoring and inspection, search-and-rescue missions, archaeological investigations, planetary explorations, etc. Even though a multiskilled single robot may accomplish several of the above-mentioned tasks, there are many applications in which, owing to the inherent characteristics of the mission/environment or a shortage of time, it is faster and cheaper to deploy a team of robots. Consider, for instance, the case of patrolling a museum: security personnel could remotely control from a central office a squad of mobile robots deployed in different zones of the museum by changing their directions, shifting their camera angles, and deciding what type of measurements to take and what mechanical movements to perform (Butler 2003). Moreover, we are able to increase the reliability and robustness of the system by combining several robots (Lima 2007). Based on the mission to be accomplished, we can also decide whether to deploy homogeneous or heterogeneous robots, i.e., robots with the same or different type of features (e.g., measuring equipment).

Regardless of their specific characteristics and tasks, being able to localize a team of mobile robots is one of the enabling features for the aforementioned applications. In particular, we emphasize that while in the case of a single robot we are interested in its *absolute* position, in the case of a team of robots we are more interested in the *relative* position of each robot with respect to its team mates. This mainly because multirobot teams usually perform actions that require a high level of coordination. As mentioned in Howard et al. (2003), "Consider, for example, a team of robots executing a formation behavior: these robots need not to know their latitude and longitude, but *must* know the relative pose of their neighbors." In this context, by exploiting the fact that mobile robots often need to communicate in order to coordinate their efforts, an approach called *cooperative localization* has been developed. In practice, each robot performs positioning measurements (e.g., radio- or image-based) with respect to its team mates via its inbuilt sensors (see Figure 9.1), and by exchanging these data over the air, more accurate location information for the whole team can be achieved (Mourikis and Roumeliotis 2006b). For a survey of related work on cooperative localization in robot networks, the reader is referred to Mourikis and Roumeliotis (2006a).

9.2.2 Wireless sensor networks

The interest in WSNs has recently grown at a tremendous pace. Like robot networks, WSNs facilitate a plethora of applications (Alhmiedat and Yang 2007): healthcare,

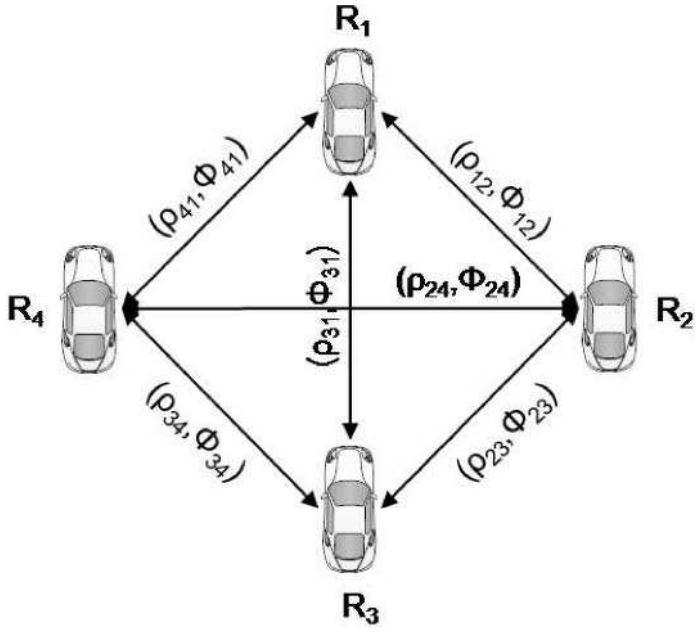


Figure 9.1 Cooperative localization in a network of mobile robots.

environmental and species monitoring, military surveillance, search and rescue, tracking soldiers and cars, etc. As mentioned in Patwari et al. (2005), "Smart structures will actively respond to earthquakes and make buildings safer; precision agriculture will reduce costs and environmental impact by watering and fertilizing only where necessary and will improve quality by monitoring storage conditions after harvesting; condition-based maintenance will direct equipment servicing exactly when and where it is needed based on data from wireless sensors; traffic monitoring systems will better control stoplights and inform motorists of alternate routes in the case of traffic jams; and environmental monitoring networks will sense air, water, and soil quality and identify the source of pollutants in real time." In general, many of these applications are enabled by the use of a large number of sensor nodes, and therefore it is highly desirable that such a network is built up in a cheap, low-energy-consuming and, as much as possible, self-organizing way. As a consequence, a sensor node – which is constituted of a battery, a sensor, a processor, a transceiver and a memory – usually has the following characteristics (Alhmiedat and Yang 2007): (i) small physical size, (ii) low power consumption, (iii) limited processing power, (iv) limited communication capabilities (only short-range) and (v) limited memory storage.

As in the case of robot networks, localization is a key feature of WSNs. Indeed, "a sensor's location must be known for its data to be meaningful" (Patwari et al. 2005). Specifically, we can classify the localization techniques envisioned for WSNs

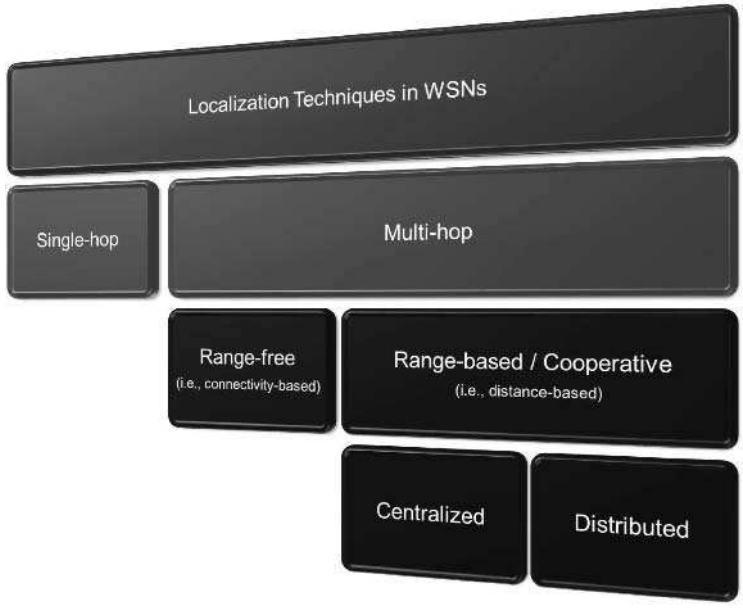


Figure 9.2 Taxonomy of localization techniques for WSNs.

as in Figure 9.2. First we distinguish between single-hop and multihop techniques. While *single-hop* refers to scenarios in which the unknown-location node is a one-hop neighbor of a sufficient number of known-location nodes (also called anchor nodes), *multihop* refers to scenarios in which the unknown-location node is $\kappa > 1$ hops from the anchor nodes. Within the realm of multihop techniques, we can divide these techniques further into range-free and range-based techniques. Unlike the former techniques, which assume that each node exploits some network-related information (e.g., connectivity information) to retrieve its own position, the latter techniques rely on distance-based measurements between nonanchor nodes also. Finally, we can differentiate range-based techniques into centralized and distributed approaches, depending on whether the information obtained by each unknown-location node is sent to a central processing node or processed independently. For a survey of work on localization in WSNs, the reader is referred to Mao et al. (2007).

In practice, single-hop techniques embrace the localization techniques described in Chapter 5. The multihop case, though, is monopolizing most of the attention, as it is the most common case in WSNs. Indeed, as mentioned before, the latter are usually vast networks and have to be kept low-cost and low-power-consuming, and therefore only very few nodes can be considered as anchor nodes (e.g., GPS-equipped nodes), and the probability for an unknown-location node to be close to them will consequently be very low. In this context, similarly to the case of relative positioning in robot networks, we can use the cooperative localization approach, where, in addition to measurements between unknown-location nodes and anchor

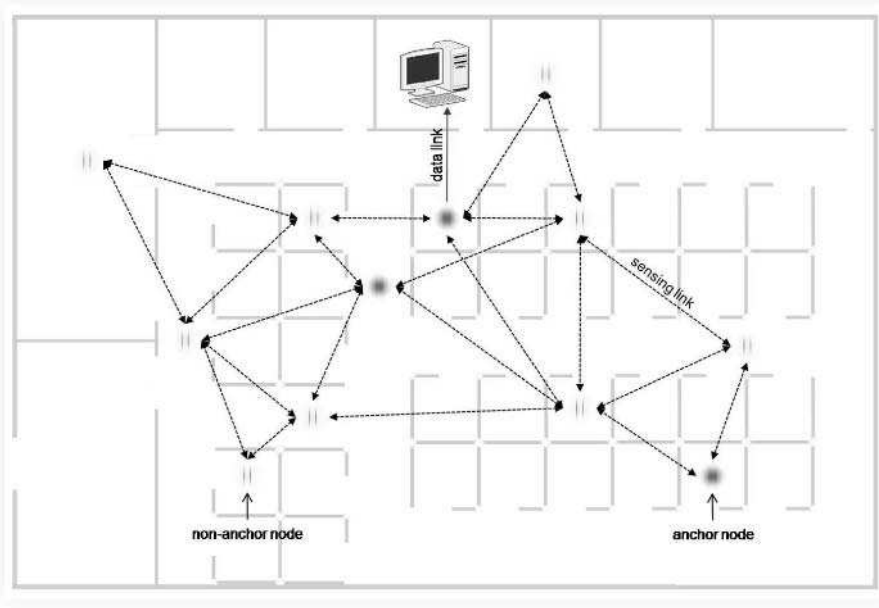


Figure 9.3 Cooperative localization in WSNs.

nodes, each unknown-location node performs local measurements with respect to its one-hop neighbors (Figure 9.3). In particular, if all this information is conveyed to a location server, the latter can calculate the estimated "absolute" positions of all the nodes in the network at once, and the accuracy and robustness of the localization system will increase (Patwari et al. 2005).

9.2.2.1 Clustering

As described above, WSNs are characterized by a high node density and the possibility of data aggregation. In particular, since the measurements performed by neighboring sensor nodes are correlated, local aggregation can be done in order to reduce the amount of data to be sent to the remote station (Mhatre and Rosenberg 2005). As a consequence, the classical operation of clustering or grouping of the sensors is performed (see Section 6.4.4.1). Practically, the space covered by the network is divided into several logical spots, inside which CHs, CMs and cluster gateways (CGs) are defined (Figure 9.4). In particular, the CH is the controller and aggregator of the group, the CMs are the sensor nodes belonging to the cluster and obeying the CH, and the CGs are the interfaces between different clusters. Besides its original purpose of lightening the process of data aggregation over a vast network, clustering also brings the following advantages (Heinzelman et al. 2002; Mhatre and Rosenberg 2004): scalability, bandwidth reuse, increased system capacity, smarter

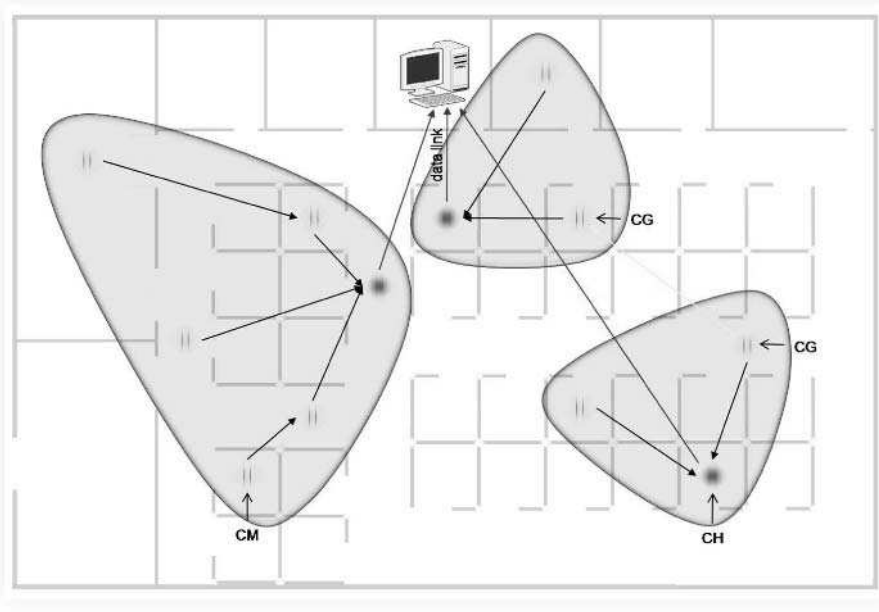


Figure 9.4 Clustering in WSNs.

resource allocation, improved power control, maximized network lifetime and fault tolerance. For a survey of work on clustering in WSNs, the reader is referred to Wei and Chan (2005).

9.2.3 Wireless mobile networks

As described in the previous sections, cooperative localization is an approach that was originally used in robot networks (for relative positioning) and later on widely adopted in WSNs (also in this case for absolute positioning, provided there is a sufficient number of anchor nodes). Such an approach has been ported recently to wireless mobile networks, i.e., networks composed of MSs immersed in a wireless or cellular layout, where the goal is to improve the accuracy of standard localization systems (satellite-based or terrestrially based) in environments where the latter do not suffice alone. In this sense, we refer to a localization system which employs the cooperative localization approach as a CAS. Practically, using the formalism of WSNs, besides having measurements between anchor nodes (i.e., BSs, APs or satellites) and unknown-location nodes (i.e., MSs) at our disposal, we rely also on measurements between unknown-location nodes. Specifically, while satellite-based and terrestrially based localization systems on their own provide high accuracy in rural and suburban scenarios, CASs are best exploitable in (1) indoor

environments, where both global navigation satellite systems (GNSSs) and cellular-based localization systems provide poor accuracy, and WLAN-based localization systems have to deal with challenging propagation conditions, and (2) outdoor environments in the presence of urban canyons, where both GNSSs and cellular-based localization systems provide poor accuracy, and a dense WLAN might not be available.

As shown in Figure 9.5, the presence of multipath effects and shadowing between the majority of MSs and the available satellites or BSs might compromise the accuracy of the localization system. In this context, the employment of the cooperative localization approach has two main benefits. (a) Owing to the short distances between users, the P2P measurements between MSs are likely to be very accurate and therefore even if all MSs are in NLOS conditions with respect to all available reference points, the accuracy of the localization system can be enhanced (Frattasi 2007; Frattasi and Figueiras 2007). (b) Owing to the spatial diversity of the MSs, if at least one of them is in LOS with respect to one reference point (e.g., MS_1 with respect to S_1 or MS_2 with respect to BS_1), we can circumvent co-located shadowing effects. Using the terminology of WSNs, the MSs in LOS can be treated as anchor nodes and the accuracy of the localization system can be reduced and equalized to the accuracy of one of those MSs over the region considered (Frattasi and Monti 2007b). Such a remarkable discovery opens up previously unimaginable scenarios, which will lead to new business models incorporating incentives to enable cooperation. For instance, we could imagine a situation where users will actually possess a virtual GPS on their mobiles, as they would be able to rely on the equipment of their neighbors. Therefore, even less sophisticated and less costly devices will be able to achieve the same accuracy level as if they had a GPS receiver embedded. On the other hand, the latter will help the former to obtain a continuous and accurate positioning service in areas where the GPS might not suffice (e.g., in outdoor urban canyons). Hence, there is in any case a benefit for all MSs (GPS-equipped and not), which might represent a sufficient incentive for utilizing the service. In addition to the previous two benefits, it has also to be considered that, as in the case of WSNs, the accuracy of a CAS increases with the density of cooperating MSs (Frattasi 2007). Therefore, since the aforementioned scenarios are the most obvious environments in which there is a high user density, one can immediately appreciable the great potential of this approach and the breakthrough that it represents which we refer to as *cooperative mobile positioning*: cooperative localization applied to wireless mobile networks.

Cooperative mobile positioning is an approach that applies naturally to heterogeneous settings where MSs embed technologies suited for both long- and short-range communications. It is possible to envisage several combinations of systems and technologies that could occur. In a macroscale indoor scenario, for example, a WLAN could be used for performing location measurements on both the AP–MS links and the MS–MS links; for instance, this could be implemented by switching the MS's Wi-Fi from infrastructure to ad hoc mode and back again

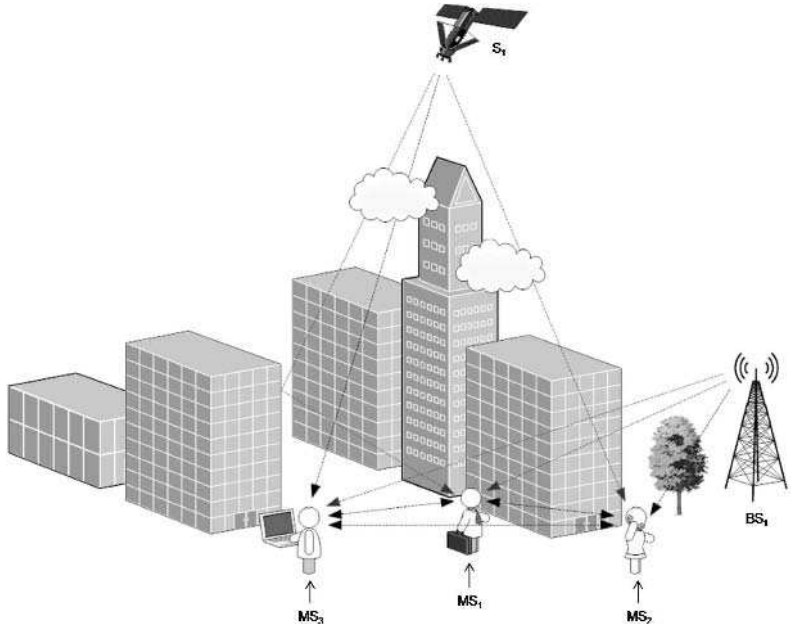


Figure 9.5 Typical "urban canyon" scenario.

(Rosa et al. 2007b). In a microscale indoor scenario, it would be better to use a WPAN (e.g., a Bluetooth-, ZigBee- or UWB-based system) to perform the short-range measurements. Finally, in an outdoor scenario, a cellular system (e.g., GSM, 3GPP-LTE or WiMAX), a GNSS (e.g., GPS) and a WLAN system could be used for performing location measurements on the BS–MS links, the satellite–MS links and the MS–MS links, respectively.

As a consequence of all of the above considerations, it clearly emerges that *cooperative mobile positioning could represent a global (indoor/outdoor) augmentation solution for the mobile localization problem*.

9.3 Cooperative Data Fusion and Filtering Techniques

In this section, we extend the algorithms shown in Chapter 5 to the cooperative framework. We also describe some specific examples of application of them, which will be extensively used in the next section.

9.3.1 Coop-WNLLS: Cooperative weighted nonlinear least squares

Given a vector function $\mathrm{f}: \mathbb{R}^q \mapsto \mathbb{R}$, a cooperative weighted nonlinear least squares (coop-WNLLS) problem is of the form

Find

$$\hat{X} = \arg\min_{X} \mathcal{J}(X),$$

where

$$\mathcal{J}(X) = \sum_{i=1}^{l} \sum_{j=1}^{m} \gamma_{i,j} [f_{i,j}(X)]^2, \tag{9.1}$$

with $l + m \geqslant q$ and $\gamma_{i,j} \geqslant 0$. For practical solutions to this minimization problem, the reader is referred to Madsen et al. (2004).

9.3.1.1 Example of application[1]

Let us consider a co-located number n_{ms} of MSs ($\mathrm{MS}_1, \ldots, \mathrm{MS}_{n_{\mathrm{ms}}}$) in range of a geographically distributed number n_{bs} of cellular BSs ($\mathrm{BS}_1, \ldots, \mathrm{BS}_{n_{\mathrm{bs}}}$). Moreover, let us assume that while each MS performs either m_{bs} TOA or TDOA measurements with respect to each available BS, only the home BS, BS_1, performs m_{bs} AOA measurements with respect to each MS. In addition, let us suppose that each MS is in range of the other MSs and performs m_{ms} RSS measurements with respect to each of them. Hence, after the estimation process, at the location server we have the following available sets of data for each MS: (i) (**TOA**, **AOA**, **RSS**) or (ii) (**TDOA**, **AOA**, **RSS**), with entries defined by Equations (5.45)–(5.47) and

$$\mathbf{RSS} = \begin{bmatrix} \mathbf{RSS}_1 \\ \vdots \\ \mathbf{RSS}_{n_{\mathrm{ms}}-1} \end{bmatrix} = \begin{bmatrix} RSS_{1,1} \cdots RSS_{1,m_{\mathrm{ms}}} \\ \vdots \\ RSS_{n_{\mathrm{ms}}-1,1} \cdots RSS_{n_{\mathrm{ms}}-1,m_{\mathrm{ms}}} \end{bmatrix}. \tag{9.2}$$

Let us write this problem in a coop-WNLLS form.

STEP 0: In addition to the definitions presented in Section 5.2.3.1, which can be simply generalized by replacing the index m with m_{bs}, we start by introducing the following definitions (Frattasi and Monti 2007b):

$$\hat{X} = \begin{bmatrix} \hat{X}^{(1)} \\ \vdots \\ \hat{X}^{(n_{\mathrm{ms}})} \end{bmatrix} = \begin{bmatrix} \hat{x}^{(1)} & \hat{y}^{(1)} \\ & \vdots \\ \hat{x}^{(n_{\mathrm{ms}})} & \hat{y}^{(n_{\mathrm{ms}})} \end{bmatrix}, \tag{9.3}$$

[1] S. Frattasi, M. Monti, "Ad-Coop positioning system (ACPS): Positioning for cooperative users in hybrid cellular ad-hoc networks", *European Transactions on Telecommunications Journal*, Wiley, vol. 19, no. 8, pp. 923–924, May, 2007. Reproduced in part by permission of © 2007 John Wiley & Sons Ltd.

where $\hat{X}$ represents the estimated location coordinates of $\mathrm{MS}_1, \ldots, \mathrm{MS}_{n_{\mathrm{ms}}}$ in a Cartesian coordinate system;

$$\hat{d}^{(i)(j)} = 10^{(\alpha - p^{(i)(j)})/10\beta}, \quad \begin{cases} i = 1, \ldots, n_{\mathrm{ms}}, \\ j = 1, \ldots, n_{\mathrm{ms}}, \\ j \neq i, \end{cases} \tag{9.4}$$

where $\hat{d}^{(i)(j)}$ is the estimated distance between MS_i and MS_j derived from $p^{(i)(j)}$, the RSS of MS_j's signal at MS_i, such that the parameters α and β are given by Equation (4.1); and

$$\hat{d}_k^{(i)(j)} = \sqrt{\left\{\hat{X}_{k-1}^{(i)} - \hat{X}_{k-1}^{(j)}\right\}^{\mathrm{T}} \left\{\hat{X}_{k-1}^{(i)} - \hat{X}_{k-1}^{(j)}\right\}}, \tag{9.5}$$

where $\hat{d}_k^{(i)(j)}$ has a similar meaning to that in Equation (9.4), but with the difference that here, instead of being calculated directly from the RSS measurements, it is derived from $\hat{X}_{k-1}^{(i)} = [\hat{x}_{k-1}^{(i)}, \hat{y}_{k-1}^{(i)}]^{\mathrm{T}}$ and $\hat{X}_{k-1}^{(j)} = [\hat{x}_{k-1}^{(j)}, \hat{y}_{k-1}^{(j)}]^{\mathrm{T}}$, the estimated location coordinates of MS_i and MS_j at iteration $k-1$ of the minimization routine.

STEP 1: The coop-WNLLS method requires initial guesses

$$\hat{X}_0 = \left[\hat{X}_0^{(1)}, \ldots, \hat{X}_0^{(n_{\mathrm{ms}})}\right]^{\mathrm{T}}$$

at the beginning of its minimization routine. These can be chosen, for example, as the outputs of an ordinary least-squares algorithm (see Section 5.2).

STEP 2: As well as Step 2 of Section 5.2.3.1, we may apply the median operator to each of the rows in Equations (5.45)–(5.47) and (9.2), and, by using the formalism introduced previously, we obtain Equations (5.55)–(5.57) and

$$\mathbf{RSS} = \begin{bmatrix} \widetilde{\mathbf{RSS}}_1 \\ \vdots \\ \widetilde{\mathbf{RSS}}_{n_{\mathrm{ms}}} \end{bmatrix} = \begin{bmatrix} p^{(1)(2)} \\ \vdots \\ p^{(1)(n_{\mathrm{ms}})} \end{bmatrix}. \tag{9.6}$$

STEP 3: Finally, we can write the objective function $\mathcal{J}(X)$ in Equation (9.1) for each of the two envisioned cases, i.e., (**TOA**, **AOA**, **RSS**) and (**TDOA**, **AOA**, **RSS**), as (Frattasi 2007)

$$\begin{aligned} \mathcal{J}(\hat{X}) = &\sum_{i=1}^{n_{\mathrm{ms}}} \sum_{j=1}^{n_{\mathrm{bs}}} \gamma_t^{(i)[j]} \{f_t(\hat{X}^{(i)})\}^2 + \sum_{i=1}^{n_{\mathrm{ms}}} \gamma_\theta^{(i)[1]} \{f_\theta(\hat{X}^{(i)})\}^2 \\ &+ \sum_{i=1}^{n_{\mathrm{ms}}} \sum_{j=1,\ j\neq i}^{n_{\mathrm{ms}}} \gamma_p^{(i)(j)} \{f_p(\hat{X}^{(i)})\}^2 \end{aligned} \tag{9.7}$$

and

$$\mathcal{J}(\hat{X}) = \sum_{i=1}^{n_{\text{ms}}} \sum_{j=1}^{n_{\text{bs}}} \gamma_{\text{t}}^{(i)[j]} \{f_{\text{t}}(\hat{X}^{(i)})\}^2 + \sum_{i=1}^{n_{\text{ms}}} \gamma_{\theta}^{(i)[1]} \{f_{\theta}(\hat{X}^{(i)})\}^2 + \sum_{i=1}^{n_{\text{ms}}} \sum_{j=1,\ j\neq i}^{n_{\text{ms}}} \gamma_{p}^{(i)(j)} \{f_{p}(\hat{X}^{(i)})\}^2, \tag{9.8}$$

where the functions f_t, f_{t} and f_θ are defined as in Equations (5.60)–(5.62) and f_p is defined as

$$f_p(\hat{X}^{(i)}) = \hat{d}^{(i)(j)} - \hat{d}_k^{(i)(j)}, \quad \begin{cases} i = 1, \ldots, n_{\text{ms}}, \\ j = 1, \ldots, n_{\text{ms}}, \\ j \neq i. \end{cases} \tag{9.9}$$

The weights γ_t, γ_{t}, γ_θ and γ_p should be appropriately selected to reflect the reliability of the available TOA, TDOA, AOA and RSS measurements. In particular, γ_t, γ_{t} and γ_θ can be chosen as in Equations (5.63)–(5.65), and γ_p as (Frattasi 2007)

$$\gamma_p^{(i)(j)} = \frac{1}{\sigma^2_{\mathbf{RSS}_j}}, \quad j = 1, \ldots, n_{\text{ms}}, \tag{9.10}$$

where $\sigma^2_{\mathbf{RSS}_j}$ is the variance directly derived from the RSS measurements available at MS_i.

N.B. The coop-WNLLS method extracts the set of location estimates which maximizes the probability that the polygon created by the conjunction of the initial position guesses has, in total, a combination of edges to the distances defined by the available sets of data. Since MS–MS links are likely to have a better channel quality than BS–MS links, the weights γ_p will have a higher value in the objective function than γ_t, γ_τ and γ_α, and thus they will tend to geometrically constrain the final solution, diminishing its location error. Therefore, it can be expected that a higher location estimation accuracy can be achieved both in LOS and in NLOS conditions compared with the noncooperative solution of Section 5.2.3.1. In particular, as mentioned in Section 9.2.3, owing to the spatial diversity of the available BSs experienced by nearby MSs, there can also be the possibility that some BS–MS links are in LOS. As a consequence, the coop-WNLLS method will treat the corresponding MSs as anchors for the others, and the remarkable result is that the location estimation accuracy can be improved and equalized over the region considered to the value for the MSs in LOS (Frattasi and Monti 2007b). However, in the case of the coop-WNLLS method when TDOA measurements are available, to declass an NLOS localization problem into an LOS localization problem is a harder matter. This is because each TDOA measurement is a function of two BSs, the home BS and a neighboring one, which is likely to be in NLOS (Cong and Zhuang 2004). Therefore, the accuracy of the AOA measurement in that case is decisive (Frattasi 2007).

9.3.2 Coop-EKF: Cooperative extended Kalman filter

In this section, we apply the EKF to the cooperative framework taking into consideration the example described in the previous section. We assume further that each MS is static, so that we can lighten the mathematical treatment of the problem (Frattasi and Figueiras 2007).

9.3.2.1 Example of application

STEP 0: In order to model the filter, it is required to first define the hidden and observable vectors as shown in Section 5.3.1. Since it is necessary to ensure cooperation among all the devices, the state space to be estimated has to include their coordinates

$$\hat{X}_k = [[\hat{X}_k^{(1)}]^{\mathrm{T}} \cdots [\hat{X}_k^{(n_{\mathrm{ms}})}]^{\mathrm{T}}]^{\mathrm{T}}, \tag{9.11}$$

where $\hat{X}_k^{(i)} = [\hat{x}_k^{(i)} \quad \hat{y}_k^{(i)}]^{\mathrm{T}}$, the superscript identifies the MS and the index k refers to a generic iteration of the estimation routine.

STEP 1: For the TOA, TDOA and RSS measurements, it is possible to define a vector of measurements for each MS:[2]

$$T_k^{(i)} = [t_k^{(i)[1]} \cdots t_k^{(i)[n_{\mathrm{bs}}]}]^{\mathrm{T}}, \tag{9.12}$$

$$\mathsf{T}_k^{(i)} = [\mathsf{t}_k^{(i)[2]} \cdots \mathsf{t}_k^{(i)[n_{\mathrm{bs}}]}]^{\mathrm{T}} \tag{9.13}$$

and

$$P_k^{(i)} = [p_k^{(i)(1)} \ \cdots \ p_k^{(i)(i-1)} \quad p_k^{(i)(i+1)} \ \cdots \ p_k^{(i)(n_{\mathrm{ms}})}]^{\mathrm{T}}, \tag{9.14}$$

where $t_k^{(i)[j]}$ is a TOA measurement between MS_i and BS_j, $\mathsf{t}_k^{(i)[j]}$ is a TDOA measurement obtained by BS_j and BS_1 with respect to MS_i, and $p_k^{(i)(j)}$ is a cooperative power measurements between MS_i and MS_j. The full set of measurements is then obtained by integrating all the measurements of the same kind into the same vector:

$$T_k = [[T_k^{(1)}]^{\mathrm{T}} \ \cdots \ [T_k^{(n_{\mathrm{ms}})}]^{\mathrm{T}}]^{\mathrm{T}}, \tag{9.15}$$

$$\mathsf{T}_k = [[\mathsf{T}_k^{(1)}]^{\mathrm{T}} \ \cdots \ [\mathsf{T}_k^{(n_{\mathrm{ms}})}]^{\mathrm{T}}]^{\mathrm{T}}, \tag{9.16}$$

$$\Theta_k = [\theta_k^{(1)[1]} \ \cdots \ \theta_k^{(n_{\mathrm{ms}})[1]}]^{\mathrm{T}} \tag{9.17}$$

and

$$P_k = [[P_k^{(1)}]^{\mathrm{T}} \ \cdots \ [P_k^{(n_{\mathrm{ms}})}]^{\mathrm{T}}]^{\mathrm{T}}, \tag{9.18}$$

for TOA, TDOA, AOA and RSS, respectively. In Equation (9.17), $\theta_k^{(i)[1]}$ represents the AOA measurement obtained by the home BS, BS_1, and MS_i. Finally, the full

[2] Since for each MS there is only a single BS that performs measurements of AOA, there is no need to specify a vector for AOA measurements.

measurements vector is defined according to the type of measurements available:

$$Z_k = [[T_k]^{\mathrm{T}} \quad [\Theta_k]^{\mathrm{T}} \quad [P_k]^{\mathrm{T}}]^{\mathrm{T}} \tag{9.19}$$

and

$$Z_k = [[\mathsf{T}_k]^{\mathrm{T}} \quad [\Theta_k]^{\mathrm{T}} \quad [P_k]^{\mathrm{T}}]^{\mathrm{T}}, \tag{9.20}$$

where (**TOA**, **AOA**) measurements and (**TDOA**, **AOA**) measurements, respectively, are available.

STEP 2: After defining the state space and the measurement space, it is necessary to specify the motion model and the observation model (see Section 5.3.1). If we assume that the MSs are static and that there is no external excitation, the new predicted state $\hat{X}_{k|k-1}$ is equal to the previous state:

$$\hat{X}_{k|k-1} = \mathcal{F}(\hat{X}_{k-1|k-1}, 0, 0) = \hat{X}_{k-1|k-1}. \tag{9.21}$$

Concerning the observation model, it is simply necessary to obtain the vector $\hat{Z}_k$ based on the models defined by Equations (5.54), (5.50), (5.51) and (9.4). In particular, we can rewrite those expressions as

$$\hat{t}^{(i)[j]} = \mathcal{H}_t(\hat{d}^{(i)[j]}) = \frac{1}{c}\hat{d}^{(i)[j]} + t_{\mathrm{tx}}, \tag{9.22}$$

$$\hat{\mathfrak{t}}^{(i)[j]} = \mathcal{H}_{\mathfrak{t}}(\hat{d}^{(i)[j]} - \hat{d}^{(i)[1]}) = \frac{1}{c}\{\hat{d}^{(i)[j]} - \hat{d}^{(i)[1]}\}, \tag{9.23}$$

$$\hat{\theta}^{(i)} = \mathcal{H}_\theta(\hat{X}^{(i)}) = \operatorname{atan}\frac{\hat{y}^{(i)} - y^{[1]}}{\hat{x}^{(i)} - x^{[1]}}, \tag{9.24}$$

$$\hat{p}^{(i)(j)} = \mathcal{H}_p(\hat{d}^{(i)(j)}) = \alpha - 10\beta \log(\hat{d}^{(i)(j)}). \tag{9.25}$$

Consequently,

$$\hat{Z}_k = \mathcal{H}_\star(\hat{X}_{k|k-1}), \tag{9.26}$$

where the entries of $\hat{Z}_k$ are calculated using $\mathcal{H}_\star(\bullet)$, which is an entrywise function that denotes the appropriate function to be used for each entry of $\hat{Z}_k$ (i.e., $\mathcal{H}_t(\bullet)$, $\mathcal{H}_\tau(\bullet)$, $\mathcal{H}_\alpha(\bullet)$ and $\mathcal{H}_p(\bullet)$).

STEP 3: Another important step in the design of the cooperative extended Kalman filter (coop-EKF) is the determination of the process and measurement noise covariance matrices. If we assume that the MSs are static, no process noise exists, and consequently Q should be equal to the zero matrix. However, in order to allow a faster convergence of the filter, a process noise adaptation method such as annealing (Merwe and Rebel 2006) can be employed, where the adaptation parameters of the method are adjusted by tuning. Concerning the measurement noise, we consider R as

a diagonal matrix (i.e., there is no correlation between measurements):

$$R = \begin{bmatrix} \sigma^2_{\text{TOA}} I & 0 & 0 \\ 0 & \sigma^2_{\text{AOA}} I & 0 \\ 0 & 0 & \sigma^2_{\text{RSS}} I \end{bmatrix} \tag{9.27}$$

and

$$R = \begin{bmatrix} \sigma^2_{\text{TDOA}} I & 0 & 0 \\ 0 & \sigma^2_{\text{AOA}} I & 0 \\ 0 & 0 & \sigma^2_{\text{RSS}} I \end{bmatrix}, \tag{9.28}$$

for (**TOA**, **AOA**) measurements and (**TDOA**, **AOA**) measurements, respectively, and I represents the identity matrix. The variances σ^2_{TOA}, σ^2_{TDOA}, σ^2_{AOA} and σ^2_{RSS} can be derived from Monte Carlo simulations. Practically, two different approaches can be used in order to set these values (Figueiras 2008):

- The position of the MS is set as constant, and the standard deviations are determined after running Monte Carlo simulations of the wireless channel conditions. This approach represents the case when a priori information concerning the noise statistics is available.

- The position of the MS is changed for every new run of the Monte Carlo simulations of the wireless channel conditions. Thus, in this case, the noise statistic are generalized and taken into account independently of the position. This approach represents the case when no a priori information concerning the noise statistics is available.

STEP 4: In order to completely model the filter, we still need to define several of the matrices in Equations (5.92) and (5.93). The matrices A_k and H_k represent the Jacobians of Equations (9.21) and (9.26), respectively, with respect to X_k. If we assume that the MSs are static, A_k is equal to the identity matrix I.

9.4 COMET: A Cooperative Mobile Positioning System

In this section, we introduce an example of a CAS, named COMET, specifically applied to cellular networks in outdoor environments (Rosa et al. 2007a). In particular, we will mainly describe two data fusion methods, which use as kernel techniques the ones illustrated in the previous section and in Chapter 5. Finally, we will reveal the potential of such a system by means of simulation results.

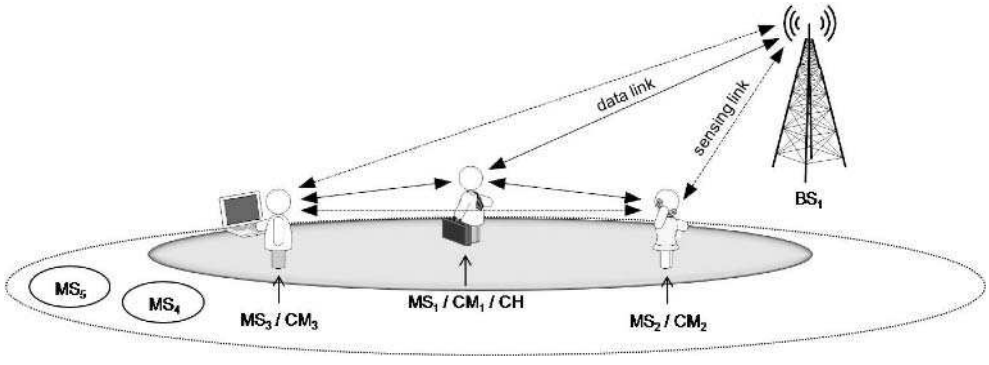


Figure 9.6 COMET: system architecture.

9.4.1 System architecture[3]

Figure 9.6 shows the hybrid cellular ad hoc system architecture taken into consideration for COMET. Upon receiving a location information request from an MS (MS_1), the home BS (BS_1) forms a cluster in the surroundings of that MS (CM_1, CM_2 and CM_3). In particular, we simply assume that the CH is the MS that sent the location request to the home BS and that the CMs are a selection of the neighboring MSs, which have been sorted in descending order according to their SNR levels on the BS–MS links. Since we will suppose in Section 9.4.3 that all MSs are in LOS with each other, we do not consider the channel conditions on the MS–MS links when performing the clustering. In general, a clustering method can also take into consideration parameters such as speed, remaining battery power and homogeneity/heterogeneity of the candidate terminals.

When the cluster has been formed, three types of measurements are carried out for location purposes. (1A) TOA measurements: each CM can measure the time of arrival of the pilot signals from the available BSs by cross-correlating each of them with an internally generated pilot signal. This type of measurement can only be used if each CM has a clock accurately synchronized with the available BSs. (1B) TDOA measurements: each CM can measure the time difference between the arrival of a pilot signal from the home BS and a neighboring BS by cross-correlating them (Knapp and Carter 1976). The pilot signal from each of the neighboring BSs can only be used if its SINR at the receiving CM is above a certain threshold. (2) AOA measurements: with an adaptive antenna array, each of the available BSs steers its antenna spot beam to track the dedicated backward-link pilot signal from each CM and thus obtains its arriving azimuth angle (with respect to a specified

[3] S. Frattasi, M. Monti, "Ad-Coop positioning system (ACPS): Positioning for cooperative users in hybrid cellular ad-hoc networks", *European Transactions on Telecommunications Journal*, Wiley, vol. 19, no. 8, pp. 923–924, May, 2007. Reproduced in part by permission of © 2007 John Wiley & Sons Ltd.

reference direction). If we consider a cellular system based on CDMA, in order to avoid signal degradation due to the near–far effect, this type of measurement can only be performed by the home BS (Cong and Zhuang 2002). (3) RSS measurements: each CM can measure the received power of the signals coming from neighboring CMs. This may be done during normal data communications without any additional bandwidth or energy requirements (Frattasi and Monti 2007a). For details on each type of measurement, the reader is referred to Chapter 4.

The determination of the location of each CM, which is performed by one of the data fusion methods that will be described in the next section, can be either *CH-based* or *CH-assisted/network-based* (the choice depends mainly on the computational power at the CH). In the first case, while each available BS collects on the backward link the AOA measurements for each BS–CM link, the CH relays back to the home BS all the TOA/TDOA and RSS measurements obtained from each BS–CM link and CM–CM link, respectively, and the calculations are performed by a specific server in the network. In the second case, instead, while the home BS transmits all the AOA measurements to the CH via the forward link, the CH collects all the TOA/TDOA and RSS measurements obtained from each BS–CM link and CM–CM link, and the calculations are performed directly at the CH. In the context of the main aim of this section, which is to show the impact of cooperative mobile positioning on the location estimation accuracy of a cellular system, the choice between CH-based and CH-assisted methods is not strictly relevant.

In this section, we consider as our hybrid cellular ad hoc system a UMTS/WLAN 802.11a system,[4] in which we make the assumption that each CM is not equipped with a GPS receiver but is equipped only with UMTS and 802.11a network interfaces, used to perform long- and short-range location measurements, respectively.

9.4.2 Data fusion methods

In this section, we describe two data fusion methods developed for COMET. While the first employs both long- and short-range location measurements together to provide absolute location coordinates for all the CMs at once, the second utilizes the measurements in parallel and may supply relative as well as absolute location information for all the CMs at once.

9.4.2.1 1L-DF: One-level data fusion[5]

Figure 9.7 shows a flow-chart of 1L-DF, which may be described in detail as follows. For each MS/CM, a first location estimate is obtained from a conventional location

[4]For related work on cooperative mobile positioning in a WiMAX/WLAN 802.11a system, the reader is referred to Rosa (2007).

[5]S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part with kind permission of © 2007 Springer Science and Business Media.

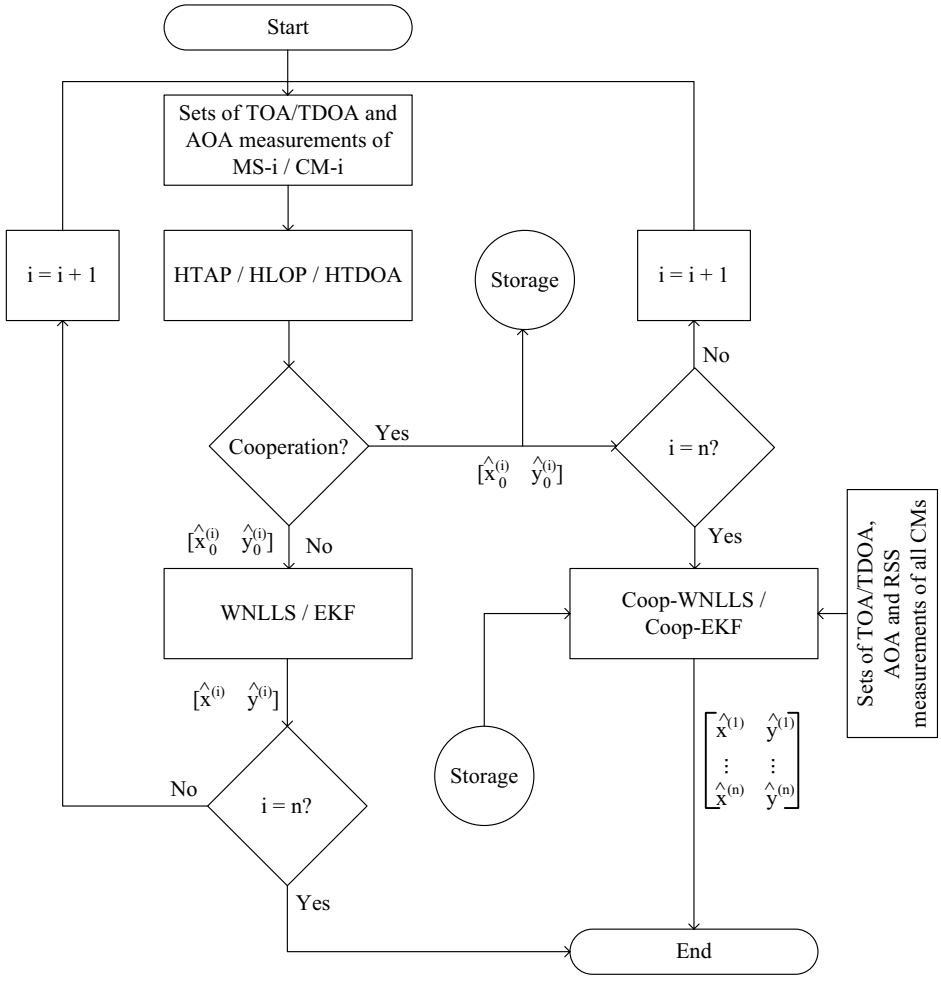


Figure 9.7 Operational representation of one-level data fusion (1L-DF).

algorithm taken from the literature. As a reference, we have considered the Hybrid TOA/AOA Positioning (HTAP) algorithm (Deng and Fan 2000) and the Hybrid Lines of Position (HLOP) algorithm (Venkatraman and Caffery 2004) for use when TOA and AOA measurements are available, and the Hybrid TDOA/AOA Positioning (HTDOA) algorithm (Chen and Feng 2005a) for use when TDOA and AOA measurements are available (see Section 4.4.5). Considering that we usually have at our disposal more than a single sample for each type of measurement, the quantities used in this step are the median values calculated over each set of measurements (for the reference statistical distribution of the error, see Section 9.4.3.1). Note that the median value is preferred to the mean value, owing to its robustness against possibly biased measurements (Frattasi and Monti 2007b). If cooperation is *off*, in

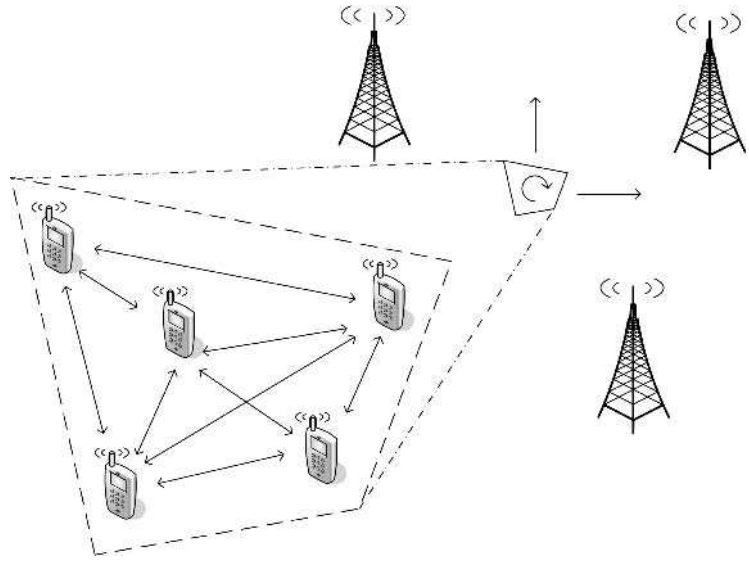


Figure 9.8 Schematic representation of the general scenario considered by the 2L-DF algorithm.

order to determine the final position estimate for each MS, each initial position guess obtained in the first step of the algorithm and each set of TOA/TDOA and AOA estimates relative to each MS are used independently in an unconstrained WNLLS minimization procedure or in an EKF (see Chapter 5). If cooperation is *on*, in order to determine the final position estimates of all the CMs at some time, all initial position guesses obtained in the first step of the algorithm and all sets of TOA/TDOA, AOA and RSS estimates relative to all the CMs are used simultaneously in an unconstrained coop-WNLLS minimization procedure or in a coop-EKF.

9.4.2.2 2L-DF: Two-level data fusion[6]

In this section, we describe two-level data fusion (2L-DF), which consists in decoupling relative and absolute localization by using independently the measurements retrieved from the short- and long-range links, respectively. As we can derive from the arrows in Figure 9.8, while relative localization is performed among the cooperating devices by estimating distances between pairs of CMs, absolute localization is performed by estimating the coordinates and orientation of a cluster within the cellular network.

[6] © 2008 IEEE. Reprinted, with permission, from Figueiras, J, Frattasi, S and Schwefel, HP, "Decoupling Estimators in Mobile Cooperative Positioning for Heterogeneous Networks" in *Proceeding of the IEEE 68th Vehicular Technology Conference*.

Like a Bayesian filter, 2L-DF runs in a cyclic way, where each iteration is carried out whenever observations are available, and performs the following consecutive steps (Figueiras 2008):

1. Decouple the absolute estimates obtained in the previous run into (i) relative coordinates and (ii) center-of-mass coordinates. In the first run, the absolute estimates are the initial guesses.
2. If measurements are available from short-range links, run a single iteration to estimate the relative coordinates. If measurements are available from long-range links, run a single iteration to estimate the center-of-mass coordinates. If measurements are available from just one domain, only the corresponding substep is executed; the advantage of this feature is that it allows short- and long-range subsystems to operate with different measurement rates.
3. Couple the newly obtained relative and center-of-mass coordinates to retrieve the current absolute estimates.

Figure 9.9 shows a flowchart of 2L-DF. In particular, it is important to notice that, if there is no request for positioning during a certain execution cycle, the coupling is not performed; this lowers the amount of computation required. Moreover, the link between the relative and absolute positioning blocks is necessary in order to define distances between CMs and BSs.

9.4.2.2.1 Decoupling and coupling The decoupling of the absolute coordinates into relative coordinates, group position and orientation is carried out by applying a transformation of coordinates. The latter is necessary to obtain the current transformation matrix (CTM), which corresponds to a translation followed by a rotation of the axis. Let us assume that the absolute coordinates of CM_i are defined by $X^{(i)}$ and its relative coordinates by $X^{(i)\mathrm{rel}}$. Moreover, let us suppose that $\mathbf{T}_{\mathrm{ctm}}$ is the CTM for the transformation of coordinates. Then, the following relation can be written:

$$\begin{bmatrix} X^{(i)\mathrm{rel}} \\ 1 \end{bmatrix} = \mathbf{T}_{\mathrm{ctm}} \begin{bmatrix} X^{(i)} \\ 1 \end{bmatrix}, \tag{9.29}$$

where the last component of the vector on the right-hand-side of Equation (9.29) can be added to allow transformations independent of $X^{(i)}$.

In order to determine $\mathbf{T}_{\mathrm{ctm}}$, a translation equivalent to the absolute position $X^{(i)}$ of CM_i is followed by a rotation equivalent to the angle of the segment between CM_j and CM_i with respect to the absolute coordinate system, $\theta = \arctan(y^{(i)(j)}/x^{(i)(j)})$. As a consequence, we can derive

$$\mathbf{T}_{\mathrm{ctm}} = \begin{bmatrix} \dfrac{x^{(i)(j)}}{d^{(i)(j)}} & -\dfrac{y^{(i)(j)}}{d^{(i)(j)}} & 0 \\ \dfrac{y^{(i)(j)}}{d^{(i)(j)}} & \dfrac{x^{(i)(j)}}{d^{(i)(j)}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -x^{(i)} \\ 0 & 1 & -y^{(i)} \\ 0 & 0 & 1 \end{bmatrix}, \tag{9.30}$$

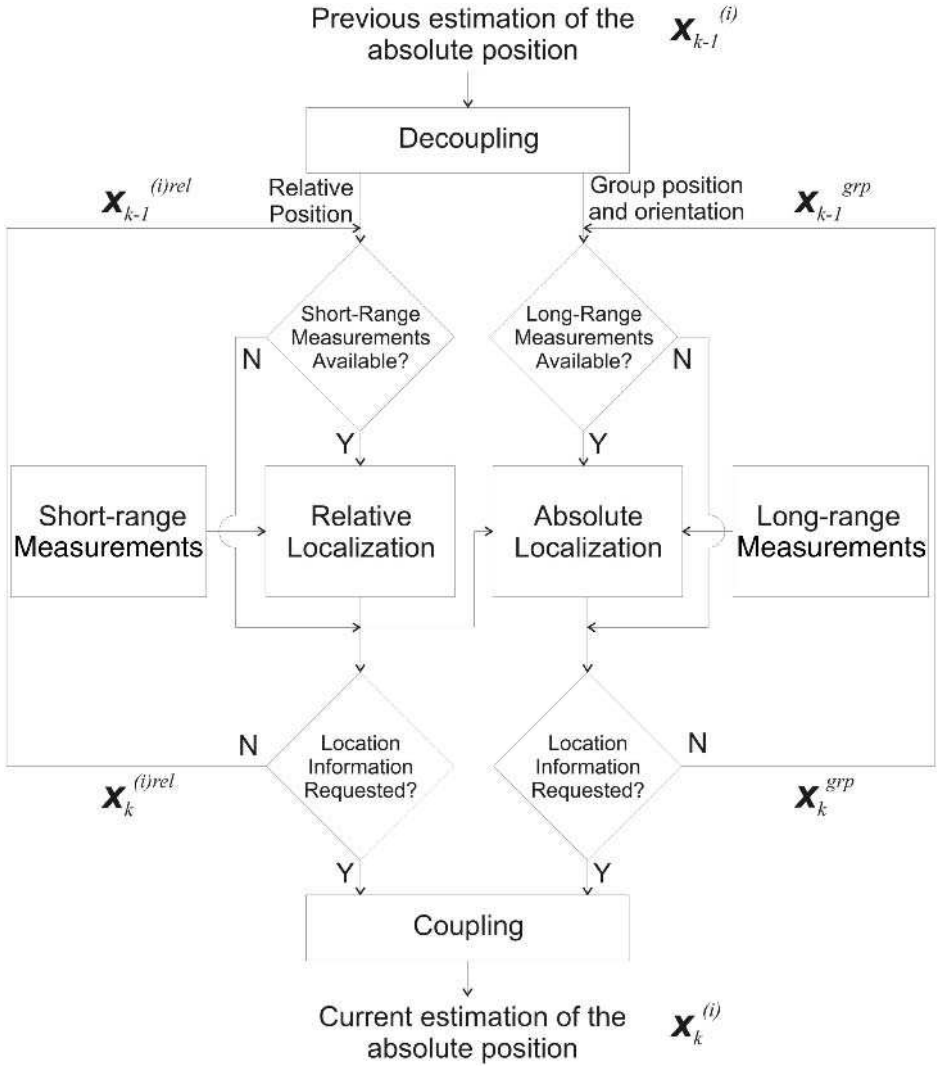


Figure 9.9 Operational representation of 2L-DF.

where $x^{(i)(j)} = x^{(j)} - x^{(i)}$, $y^{(i)(j)} = y^{(j)} - y^{(i)}$ and $d^{(i)(j)}$ is the Euclidean distance between CM_i and CM_j. Note that in this transformation of coordinates, CM_i is considered as the reference CM, i.e., the CH, and CM_j is assumed to be on the x-axis of the relative coordinate system.

As is noticeable from Figure 9.9, the decoupling implies the calculation of (i) the relative locations of the users ($X^{(i)\mathrm{rel}}$, with $i = 1, \ldots, n$) and (ii) the position and orientation of the group (X^{grp}). The relative coordinates of the CMs can be obtained by using Equation (9.29) with the CTM of Equation (9.30). The coordinates

and orientation of the group are obtained as $X^{(i)}$ and $\theta = \arctan(y^{(i)(j)}/x^{(i)(j)})$, respectively.

When the coupling is performed, the absolute coordinates of the CMs are obtained by including the coordinates and orientation of the group in Equation (9.30) and applying the inverse calculation of Equation (9.29).

9.4.2.2.2 Relative localization To perform an estimation of the relative localization of the CMs, an EKF is used (see Section 5). This choice depends on the fact that the system has a nonlinear nature that is not very accentuated, i.e., the linearization by itself does not introduce considerable error when compared with the noise introduced by the processes involved.

In order to design the EKF, one has first to define the hidden process and the observable process. In the present case, the hidden process is the set of relative positions, and the observations are the measured power values between the CMs:

$$X^{\text{rel}} = [x^{(2)\text{rel}} \quad \{X^{(3)\text{rel}}\}^{\text{T}} \quad \cdots \quad \{X^{(n_{\text{ms}})\text{rel}}\}^{\text{T}}]^{\text{T}}, \tag{9.31}$$

$$Z^{\text{rel}} = \begin{bmatrix} p^{(1)(2)}, & p^{(1)(3)}, & \cdots & p^{(1)(n_{\text{ms}})}, \\ & p^{(2)(3)}, & \cdots & p^{(2)(n_{\text{ms}})}, \\ & & \ddots & \vdots \\ & & & p^{(n_{\text{ms}}-1)(n_{\text{ms}})} \end{bmatrix}^{\text{T}}. \tag{9.32}$$

Note that Equation (9.31) does not include the coordinates of CM_1 and the y coordinate of CM_2, owing to the fact that they are equal to zero in the relative coordinate system; in order to reduce the state space, those zeros were excluded. It is also important to notice that Z^{rel} is a column vector, as we assume that short-range links are symmetric, i.e., for each pair of CMs i and j in Equation (9.32) there is a single observation $p^{(i)(j)}$, with $i < j$. If the links were not symmetric, Equation (9.32) would need to include $p^{(i)(j)}$ with $i \neq j$.

The second step in the design of the filter is the definition of the motion model and the observation model. Since we will assume static devices in the next section, the predicted state is simply equal to the previous estimation, while the observation model is defined by the path loss equation:

$$\hat{X}^{\text{rel}}_{k|k-1} = \hat{X}^{\text{rel}}_{k-1}, \tag{9.33}$$

$$\hat{p}^{(i)(j)\text{rel}}_{k} = \alpha - 10\beta \log(\hat{d}^{(i)(j)}), \tag{9.34}$$

where the subscript $k|k-1$ refers to a prediction at time k given observations up to time $k-1$. In Equation (9.34), α and β correspond to the parameters in Equation (4.1).

The final design issue is the definition of the process noise Q and measurement noise R. It is assumed that the noise components are uncorrelated, i.e., the covariance

error matrix Q and the observation covariance error matrix R are diagonal. Concerning the measurement noise, the standard deviation of the measurements σ_p^2 is used for the entries of R:

$$R = \sigma_p^2 I, \tag{9.35}$$

where the matrix I is the identity matrix with the same dimension as the observation vector in Equation (9.32). Although σ_p^2 is defined based on the observations, its value could instead be obtained either from the literature, computer simulations or a precalibration phase. Concerning the process noise Q, we can assume it to be zero, since we suppose that we have static devices. However, this approach could imply a need for a large number of measurements to allow us to ignore the additional errors introduced into the position estimation by the slow convergence of the algorithm. For this reason, a process noise adaptation method (e.g., annealing) is used to allow a faster convergence of the filter, where the adaptation parameters of that method are adjusted by tuning.

9.4.2.2.3 Absolute localization In contrast to relative localization, which is performed by means of power measurements, absolute localization shows a strongly nonlinear behavior owing to the observation model for angle-of-arrival measurements. Since the latter are related to the position coordinates by an arctan function, the linearization turns out to introduce additional errors into the process estimation. For this reason, a UKF is used, as it is expected to perform better than an EKF (see Section 5.3.1).

In order to design the UKF, one has first to define the hidden process and the observable process. In the present case, the hidden process is the set of absolute positions, and the observations are the measured time and angle values between the CMs and the BSs:

$$X^{\text{grp}} = \left[x^{(1)} \quad y^{(1)} \quad \arctan\left(\frac{y^{(1)(2)}}{x^{(1)(2)}}\right) \right]^{\text{T}}, \tag{9.36}$$

$$Z^{\text{grp}} = \left[\{T^{(1)}\}^{\text{T}} \quad \{T^{(n_{\text{ms}})}\}^{\text{T}} \quad \theta^{(1)[1]} \quad \cdots \quad \theta^{(n_{\text{ms}})[1]} \right]^{\text{T}}, \tag{9.37}$$

$$T^{(i)} = [t^{(i)[1]} \quad \cdots \quad t^{(i)[n_{\text{bs}}]}]^{\text{T}}, \tag{9.38}$$

where n_{bs} is the number of BSs and n_{ms} is the number of MSs.

The second step in the design of the filter is the definition of the motion model and the observation model. Since we will assume static devices in the next section, the predicted state is simply equal to the previous estimation, while the observation model is given by three different expressions, one for the angle measurements and

two for the time measurements, for the TOA and TDOA:

$$\hat{X}^{\text{grp}}_{k|k-1} = \hat{X}^{\text{grp}}_{k-1}, \tag{9.39}$$

$$\theta^{(i)[1]} = \arctan\left(\frac{\hat{y}^{(i)} - y^{[1]}}{\hat{x}^{(i)} - x^{[1]}}\right), \tag{9.40}$$

$$t^{(i)[1]} = cd^{(i)[1]}, \tag{9.41}$$

$$\mathfrak{t}^{(i)[j]} = c(d^{(i)[j]} - d^{(i)[1]}), \tag{9.42}$$

where c is the speed of light and $d^{(i)[j]}$ is the distance between CM_i and BS_j. It is important to notice that the position of CM_i is obtained from X^{grp} and the relative distances between the CMs, assuming that those distances are fixed parameters. This step corresponds to the link between the relative and absolute localization blocks in Figure 9.9.

The final design issue is the definition of the process noise Q and measurement noise R. As in Section 9.4.2.2.2, it is assumed that the noise components are uncorrelated, i.e., the covariance error matrix Q and the observation covariance error matrix R are diagonal. Concerning the measurement noise, the standard deviation of the time measurements σ_t^2 and angle measurements σ_θ^2 are used in the entries of R:

$$R = \left[\begin{array}{c|c} \sigma_t^2 I & 0 \\ \hline 0 & \sigma_\theta^2 I \end{array}\right], \tag{9.43}$$

where the matrix I is the identity matrix with the same dimension as the observation vector in Equation (9.39). Although σ_t^2 and σ_θ^2 are defined based on the observations, their values could instead be obtained from either the literature, computer simulations or a precalibration phase. Concerning the process noise, Q is modeled as described for the case of relative localization.

9.4.2.2.4 Properties of 2L-DF The decoupling and coupling of the position estimators in 2L-DF has the following advantages:

- **Different observation rates.** Positioning solutions generally operate with different measurement rates depending on the technology in use. Since it is common that long- and short-range communications are performed by different technologies, their integration in terms of positioning data requires attention. The problem is addressed by the decoupling of the position estimators, which allows the two domains of data sources to operate independently (Figure 9.9).

- **Computational effort and distributed computing.** Given the structure of 2L-DF, the computational effort is at most as large as that of a data fusion where all measurements are treated equally. Additionally, owing to the decoupling of the estimators, distributed computation is possible to a certain extent. For instance, relative localization could be carried out by the CMs, while absolute

localization could be carried out by the cellular network. Then, depending on where the information is requested, the necessary exchange of data could be performed.

- **Group mobility.** The decoupling of short- and long-range measurements permits one to use group mobility models (see Chapter 6) instead of individual mobility models. In this way, the CMs can be viewed, on the one hand, as a group with correlated mobility patterns, and, on the other hand, as individuals characterized by some uncorrelated mobility within the group. A consequence of such an approach is that management of groups is required when devices move out of range.

9.4.3 Performance evaluation

9.4.3.1 Simulation models[7]

Below, we list the models used to obtain the simulation results illustrated in the next section. The statistical models for the TOA and AOA estimation errors were taken from Deng and Fan (2000), where the differentiation between LOS and NLOS is based on results concerning the LOS probability from Wang et al. (2004), from which the channel model used for the estimation of the RSS was also extracted. In particular, in Wang et al. (2004), a statistical channel model was derived from a ray-tracing simulation of the city of Bristol for both BS–MS and MS–MS links at frequencies of 2.1 GHz (UMTS) and 5.2 GHz (WLAN 802.11a), respectively.

9.4.3.1.1 Statistical models for time and angle-of-arrival estimation errors[8]

- **Line-of-sight.** The estimation errors are small and are primarily due to equipment measurement errors. Traditionally, they have been assumed to be normally distributed with zero mean and small standard deviation: $c\sigma_T = 30$ m and $\sigma_\Theta = 1$ deg for TOAs and AOAs, respectively (Deng and Fan 2000).[9] Note that we suppose that the standard deviations for TOA measurement errors associated with different BSs are identical.

- **Non-line-of-sight.** The estimation errors are large and are primarily due to reflection or diffraction of the signal between a BS and a CM. Consequently,

[7] S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part with kind permission of © 2007 Springer Science and Business Media.

[8] S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part with kind permission of © 2007 Springer Science and Business Media.

[9] In practice, the standard deviation of the measurement errors depends on the chip rate, the propagation environment and the home BS antenna parameters. It can be estimated based on the SNR and its typical values for various propagation conditions (Cong and Zhuang 2004).

Table 9.1 Parameter settings for different environmental types (Greenstein et al. 1997). (S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part by kind permission of © Springer Science and Business Media.)

Channel type	$\widetilde{\tau_{\text{rms}}}$ [μs]	ε	σ_χ (dB)
(A) Urban macrocells	0.4–1.0	0.5	1.9–3.6
(B) Urban microcells	0.4	0.5	2.3
(C) Suburban areas	0.3	0.5	2.0–4.7
(D) Rural areas	0.1	0.5	4.0–5.3
(E) Mountainous areas	≥ 0.5	1.0	2.4–3.2

the following exponential probability density function can be used to model the excess delay (Greenstein et al. 1997):

$$\mathrm{P}(\tau) = \frac{1}{\tau_{\text{rms}}} \exp -\tau/\tau_{\text{rms}}, \tag{9.44}$$

where τ_{rms} is the RMS delay spread:

$$\tau_{\text{rms}} = \widetilde{\tau_{\text{rms}}}\{\hat{d}^{(i)[j]}\}^{\varepsilon}\chi, \tag{9.45}$$

where $\widetilde{\tau_{\text{rms}}}$ is the median value of the RMS delay spread in μs at 1 km, $\hat{d}^{(i)[j]}$ is the distance between BS_j and CM_i in km, ε is an exponent with a value between 0.5 and 1, and χ represents the log-normal shadow fading, so that $X = 10 \log \chi$ is a Gaussian random variable with zero mean $\mu_\chi = 0$ and a standard deviation σ_χ that lies between 2 and 6 dB (Table 9.1). Specifically, the autocorrelation of the shadow fading among the CMs is modeled by (Senarath 2006)

$$\rho(\Delta x) = e^{-(|\Delta x|/d_{\text{cor}})\ln 2}, \tag{9.46}$$

where ρ is the correlation coefficient, Δx is the distance between the CH and a given CM, and $d_{\text{cor}} = 20$ m is the decorrelation distance. For example, if χ_1 is the log-normal component at position P_1, the component at position P_2, χ_2, which is Δx away from P_1, is normally distributed with mean $\mu_2 = \rho(\Delta x)\chi_1$ dB and standard deviation $\sigma_2 = \sigma_\chi\sqrt{1 - [\rho(\Delta x)]^2}$ dB.

The estimation error in the angle measurement is considered to be Gaussian distributed with zero mean $\mu_\Theta = 0$ and a standard deviation given by (Deng and Fan 2000)

$$\sigma_\theta = \frac{c\tau}{d^{(i)[j]}}, \tag{9.47}$$

where τ is the excess delay determined by use of Equation (9.44).

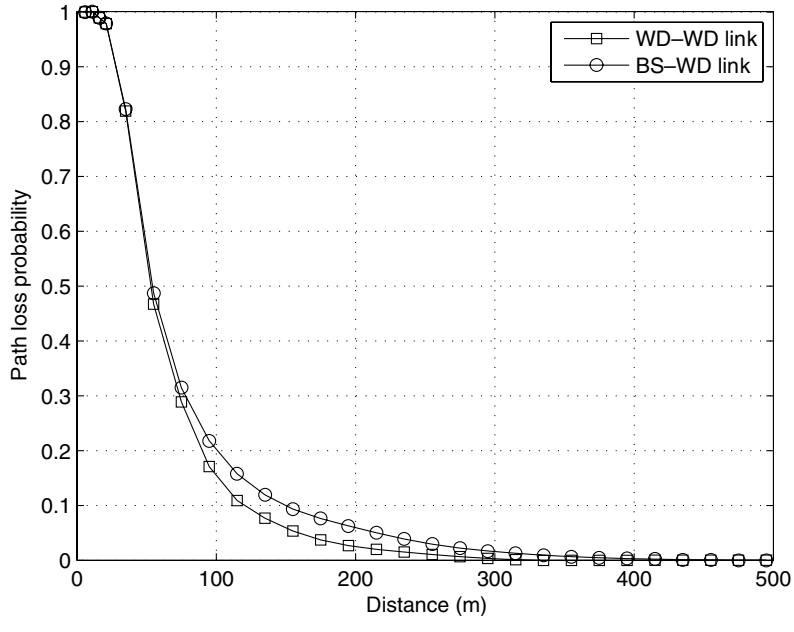


Figure 9.10 LOS probability vs. separation distance between TX and RX (Frattasi and Monti 2007a).

9.4.3.1.2 Statistical channel model for received-signal-strength estimation[10]

- **Line-of-sight probability.** Figure 9.10 shows the LOS probability as a function of the separation distance between the transmitter (TX) and receiver (RX), for both BS–CM and CM–CM links. The following features can be observed. (1) When the separation distance between TX and RX lies in the region of 10–20 m, which broadly corresponds to the width of a street, the LOS probability is one, for this particular database and set of locations (Wang et al. 2004). (2) When the separation distance between TX and RX increases further, the LOS probability is higher for BS–CM links than for CM–CM links. This is because when the height of TX is reduced from 15 m for BS–CM links to 1.5 m for CM–CM links, the signal propagation paths are affected not only by the surrounding buildings in the vicinity of TX but also by the terrain. As a consequence, CM–CM links can also be in NLOS even when no buildings lie in the propagation path.

[10] S. Frattasi, M. Monti, "Ad-Coop positioning system (ACPS): Positioning for cooperative users in hybrid cellular ad-hoc networks", *European Transactions on Telecommunications Journal*, Wiley, vol. 19, no. 8, pp. 923–924, May, 2007. Reproduced in part by permission of © 2007 John Wiley & Sons Ltd.

- **Path loss.** The following slope/intercept statistical model is used for the mean outdoor path loss:

$$L = L_0 + 20\ \log_{10}(f) + 10\beta\ \log_{10}(d), \tag{9.48}$$

where d is the separation distance between TX and RX in m, f is the operating frequency in MHz and L_0 is the path loss in dB at a short reference distance $d_0 = 1$ m. For CM–CM links, L is modeled in the LOS case with $L_0 = -27.6$ dB and $\beta = 2$ (free-space path loss), and in the NLOS case with $L_0 = -51.22$ dB and $\beta = 5.82$.

- **Shadowing.** The shadowing process χ is characterized by a log-normal distribution, i.e., a normal distribution in dB, with a distance-dependent standard deviation given by

$$\sigma_\chi = s[1 - e^{-(d-d_0)/D_s}], \tag{9.49}$$

where s is the maximum standard deviation in dB and D_s is the growth distance factor in m. For CM–CM links, σ_χ is modeled in the LOS case with $s = 2$ dB, $D_s = 36$ m and $d_0 = 0$ m, and in the NLOS case with $s = 23.4$ dB, $D_s = 36$ m and $d_0 = 10$ m.

- **Fast fading.** The most popular models for fast fading in LOS and NLOS are, respectively, the Rice and Rayleigh models, where the ratio between the expected power of the dominant path ρ^2 and the power of the Rayleigh components $2\sigma^2$ is often expressed by the *Ricean K-factor* $\kappa_0 = \rho^2/2\sigma^2$. For CM–CM links, κ_0 is modeled in the LOS case by

$$\kappa_0 = -n_k d + L_k, \tag{9.50}$$

with $L_k = 23$ dB and $n_k = 0.029$, and in the NLOS case by

$$\kappa_0 = \log \text{Norm}(\mu_k, \sigma_k) - 10, \tag{9.51}$$

where $\log \text{Norm}(\mu_k, \sigma_k)$ is a log-normal distribution with mean $\mu_k = 2.43$ dB and standard deviation $\sigma_k = 0.45$ dB.

9.4.3.2 Simulation results

Computer simulations were performed in order to compare the location estimation accuracy of COMET with that of conventional hybrid positioning techniques in stand-alone cellular networks. Simulated urban microcells of radius 1000 m were considered, where a hexagonal test cell is surrounded by six neighboring cells (Figure 9.11). Specifically, $1 \leq N \leq 7$ BSs were positioned in a two-dimensional plane with location coordinates BS_1 (0, 0) m, BS_2 (1732, 1000) m, BS_3 (1732, −1000) m, BS_4 (0, 2000) m, BS_5 (0, −2000) m, BS_6 (−1732, 1000) m

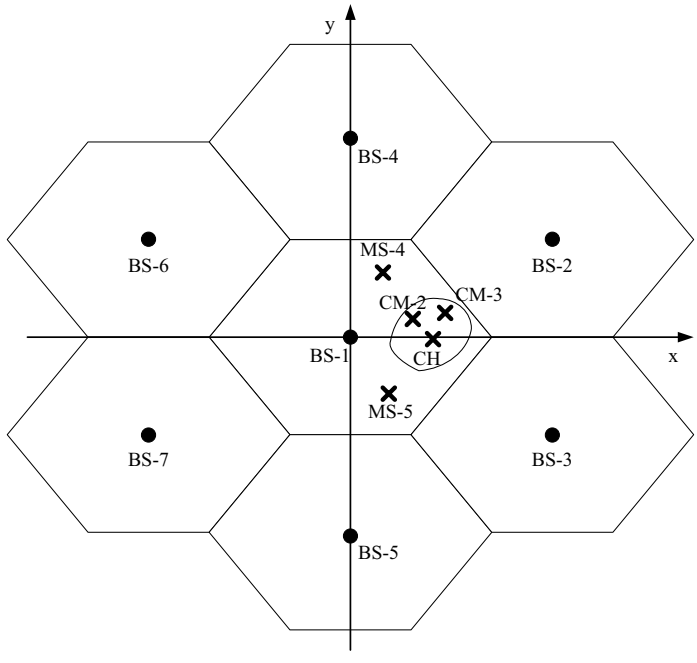


Figure 9.11 The seven-cell and one-cluster system layout considered (Frattasi et al. 2006).

and BS_7 $(-1732, -1000)$ m. For simplicity, we took into account only one cluster of radius 25 m ($P_{\text{los}} = 1$ for CM–CM links), which embraced $1 \leq n \leq 8$ CMs, where the CH was placed in the center and the other CMs were uniformly generated around it. In Sections 9.4.3.2.1 and 9.4.3.2.2, only n MSs are generated around the CH and considered as CMs. In Section 9.4.3.2.3, we show the effect of clustering by generating up to 25 MSs and selecting n of those according to their SNR levels with respect to the available BSs (see earlier in Section 9.4). Finally, no interference was considered, and 1L-DF with coop-WNLLS was used as a data fusion method; however, similar results can be obtained with 1L-DF and the coop-EKF or with 2L-DF (Figueiras 2008). The values of some of the parameters used in the simulations are summarized in Table 9.2.

The performance of COMET was presented in the form of cumulative distribution functions (CDFs) of the cooperative root mean square error (CRMSE) (Frattasi and Monti 2006):

$$\text{CRMSE} = E[\mathbf{RMSE}], \tag{9.52}$$

where $E[\bullet]$ is the operator "mean" and $\mathbf{RMSE} = \left[RMSE_1 \cdots RMSE_n\right]^{\text{T}}$ is the RMSEs for all the CMs, which are defined as

$$RMSE_i = \sqrt{\{\hat{X}^{(i)} - X_{\text{r}}^{(i)}\}^{\text{T}}\{\hat{X}^{(i)} - X_{\text{r}}^{(i)}\}}, \tag{9.53}$$

Table 9.2 Simulation parameters. (S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part by kind permission of © Springer Science and Business Media.)

Parameter	Value
Cell radius	1000 m
Cluster radius	25 m
Number of BSs	$N = 1$–7
Number of CMs	$n = 1$–8
Measurements per BS	$I = 100$
Measurements per CM	$i = 100$
Channel type (see Table 9.1)	(B)

where $X_{\mathrm{r}}^{(i)}$ is the vector representing the real location coordinates of CM_i. If cooperation is *off*, the CRMSE becomes equal to the RMSE, which is calculated by considering only one MS, the CH. From Equation (9.52), we can observe further that if cooperation is *on*, the accuracy experienced by a single MS has actually been replaced with the accuracy experienced by all the CMs.

9.4.3.2.1 Dependence of performance on the number of cluster members[11]

- **HTAP.** Figure 9.12 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF with and without cooperation for a variable number of CMs, when the CH is placed at (500, 0) m ($P_{\mathrm{los}} = 0$ for BS–CM links) and the initial guesses are calculated by HTAP. The following can be observed. (1) Regardless of the number of CMs involved, the case with cooperation on always outperforms the case with cooperation off. In particular, for a setup with $n = 8$, the location estimation accuracy is increased on average by about 48% (Table 9.3). (2) The CRMSE drops considerably when the number of CMs is increased. This is because the location estimates obtained in the first step of the data fusion method must represent a polygonal configuration in space that respects the geometrical constraints imposed by the knowledge about the relative distances between the CMs. As a consequence,

[11] S. Frattasi, M. Monti, "Ad-Coop positioning system (ACPS): Positioning for cooperative users in hybrid cellular ad-hoc networks", *European Transactions on Telecommunications Journal*, Wiley, vol. 19, no. 8, pp. 923–924, May, 2007 and S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightement of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part by permission of © 2007 John Wiley & Sons Ltd. © 2007 Springer Science and Business Media.

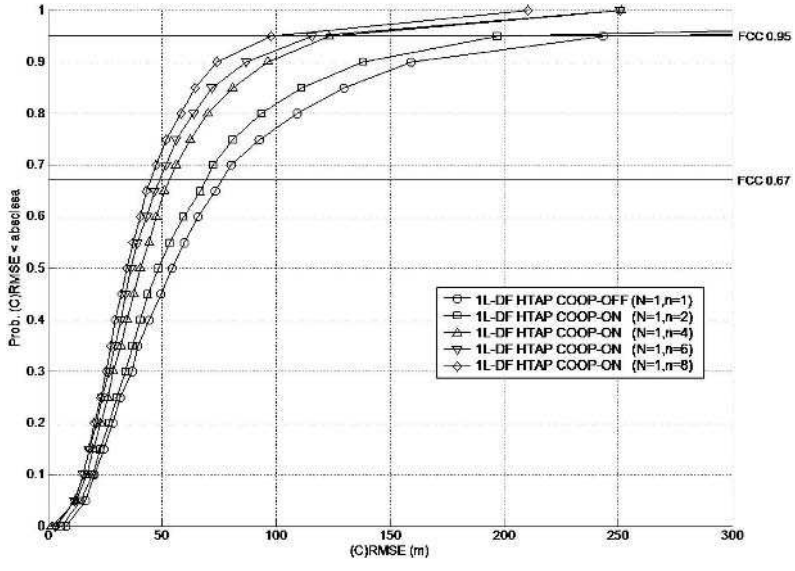


Figure 9.12 (C)RMSE vs. number of CMs (HTAP case). (S. Frattasi, M. Monti, "Cooperative Mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions, Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part by kind permission of © Springer Science and Business Media.)

the more cooperative users join the cluster, the more constraints the selected solution has to respect and the more accurate the final location estimation is. For example, on going from $n = 2$ to $n = 8$, the location estimation accuracy is increased on average by about 40% (Table 9.3). (3) For the specific simulation models considered here, both cases (cooperation on and cooperation off) always meet the FCC requirements.

- **HLOP.** Figure 9.13 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF with and without cooperation for a variable number of CMs, when the CH is placed at (500, 0) m ($P_{\text{los}} = 0$ for BS–CM links) and the initial guesses are calculated by HLOP. The following can be observed. (1) Regardless of the number of CMs involved, the case with cooperation on always outperforms the case with cooperation off. In particular, for a setup with $n = 8$, the location estimation accuracy is increased on average by about 61% (Table 9.3). (2) The CRMSE drops considerably when the number of CMs is increased. For example, on going from $n = 2$ to $n = 8$, the location estimation accuracy is increased on average by about 49% (Table 9.3). (3) 1L-DF with HLOP always performs better than 1L-DF with HTAP and HTDOA (see the next section). For example, for a setup with $n = 8$, the location estimation accuracy is increased on average by about 30% and 7%

Table 9.3 (C)RMSE statistics for a variable number of CMs and BSs

Data fusion algorithm	Cooperation	N	$n = 1$	$n = 2$	$n = 4$	$n = 6$	$n = 8$	Metrics
HTAP-WNLLS	Off	1	80.66, 91.85	–, –	–, –	–, –	–, –	$\begin{bmatrix} \mu & , & \sigma \\ 67\% & , & 95\% \end{bmatrix}$ (m, m)
			76.07, 244.4	–, –	–, –	–, –	–, –	
	On		–, –	69.78, 71.45	50.53, 37.22	46.51, 35.65	41.79, 28.50	
			–, –	68.79, 207.1	52.35, 125.0	49.30, 116.1	44.60, 98.51	
HLOP-WNLLS	Off	3	74.18, 52.54	–, –	–, –	–, –	–, –	
			84.94, 167.8	–, –	–, –	–, –	–, –	
	On		–, –	57.67, 46.04	36.03, 29.18	32.64, 25.37	29.16, 23.37	
			–, –	66.49, 144.2	42.79, 93.89	39.22, 80.48	33.74, 77.56	
HLOP-WNLLS	Off	5	38.01, 24.36	–, –	–, –	–, –	–, –	
			42.91, 83.48	–, –	–, –	–, –	–, –	
	On		–, –	30.59, 21.44	20.98, 14.30	18.77, 12.83	17.99, 12.18	
			–, –	33.13, 71.53	23.70, 48.92	21.30, 43.87	19.85, 41.51	
HTDOA-WNLLS	Off	3	81.37, 51.02	–, –	–, –	–, –	–, –	
			97.72, 151.08	–, –	–, –	–, –	–, –	
	On		–, –	67.79, 44.64	42.09, 29.12	37.67, 25.54	31.31, 20.76	
			–, –	83.72, 138.34	50.77, 94.24	45.87, 80.58	35.94, 70.45	
HTDOA-WNLLS	Off	5	55.42, 51.78	–, –	–, –	–, –	–, –	
			58.42, 124.81	–, –	–, –	–, –	–, –	
	On		–, –	44.91, 38.85	32.34, 27.05	29.28, 22.51	27.29, 20.59	
			–, –	48.81, 107.08	35.67, 77.32	32.12, 66.24	30.80, 59.95	

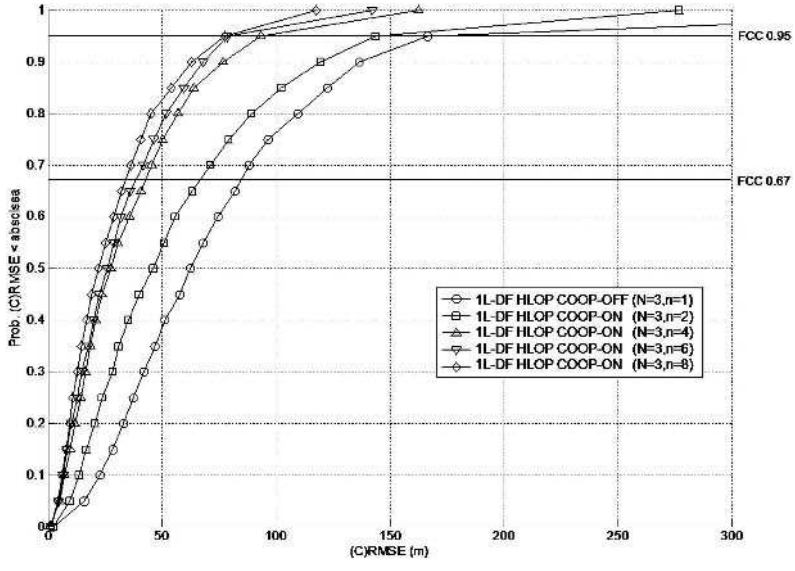


Figure 9.13 (C)RMSE vs. number of CMs (HLOP case). (S. Frattasi, M. Monti, "Cooperative Mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions, Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part by kind permission of © Springer Science and Business Media.)

with respect to 1L-DF with HTAP and with HTDOA, respectively (Table 9.3). (4) For the specific simulation models considered here, both cases (cooperation on and cooperation off) always meet the FCC requirements.

- **HTDOA.** Figure 9.14 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF with and without cooperation for a variable number of CMs, when the CH is placed at (500, 0) m ($P_{\text{los}} = 0$ for BS–CM links) and the initial guesses are calculated by HTDOA. The following can be observed. (1) Regardless of the number of CMs involved, the case with cooperation on always outperforms the case with cooperation off. In particular, for a setup with $n = 8$, the location estimation accuracy is increased on average by about 61% (Table 9.3). (2) The CRMSE drops considerably when the number of CMs is increased. For example, on going from $n = 2$ to $n = 8$, the location estimation accuracy is increased on average by about 54% (Table 9.3). (3) For the specific simulation models considered here, both cases (cooperation on and cooperation off) always meet the FCC requirements.

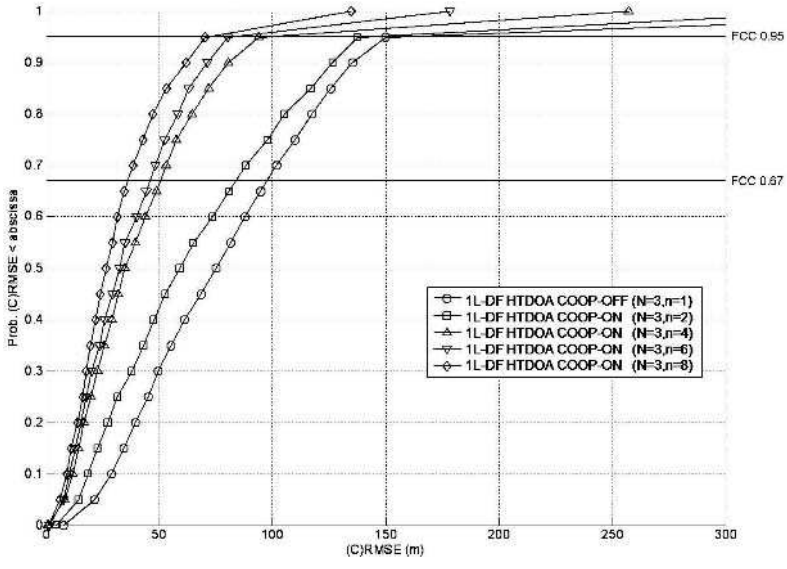


Figure 9.14 (C)RMSE vs. number of CMs (HTDOA case).

9.4.3.2.2 Dependence of performance on the number of base stations[12]

- **HLOP.** Figure 9.15 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF with and without cooperation for a variable number of BSs, when the CH is placed at (500, 0) m ($P_{\text{los}} = 0$ for BS–CM links) and the initial guesses are calculated by HLOP. The following can be observed. (1) The CRMSE drops considerably when the numbers of BSs and CMs are increased. For example, on going from $[N = 3, n = 8]$ to $[N = 5, n = 8]$, the location estimation accuracy is increased on average by about 38% (Table 9.3). (3) For the specific simulation models considered here, COMET achieves an accuracy very much comparable to that of the GPS for the combination $[N = 5, n = 8]$.

- **HTDOA.** Figure 9.16 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF and without cooperation for a variable number of BSs, when the CH is placed at (500, 0) m ($P_{\text{los}} = 0$ for BS–CM links) and the initial guesses are calculated by HTDOA. The following can be observed. (1) The CRMSE drops considerably when the numbers of BSs and CMs are increased. For example, on going from $[N = 3, n = 8]$ to

[12] S. Frattasi, M. Monti, "Cooperative mobile positioning in 4G wireless networks", *Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications*, Springer, pp. 213–233, September, 2007. Reproduced in part with kind permission of © 2007 Springer Science and Business Media.

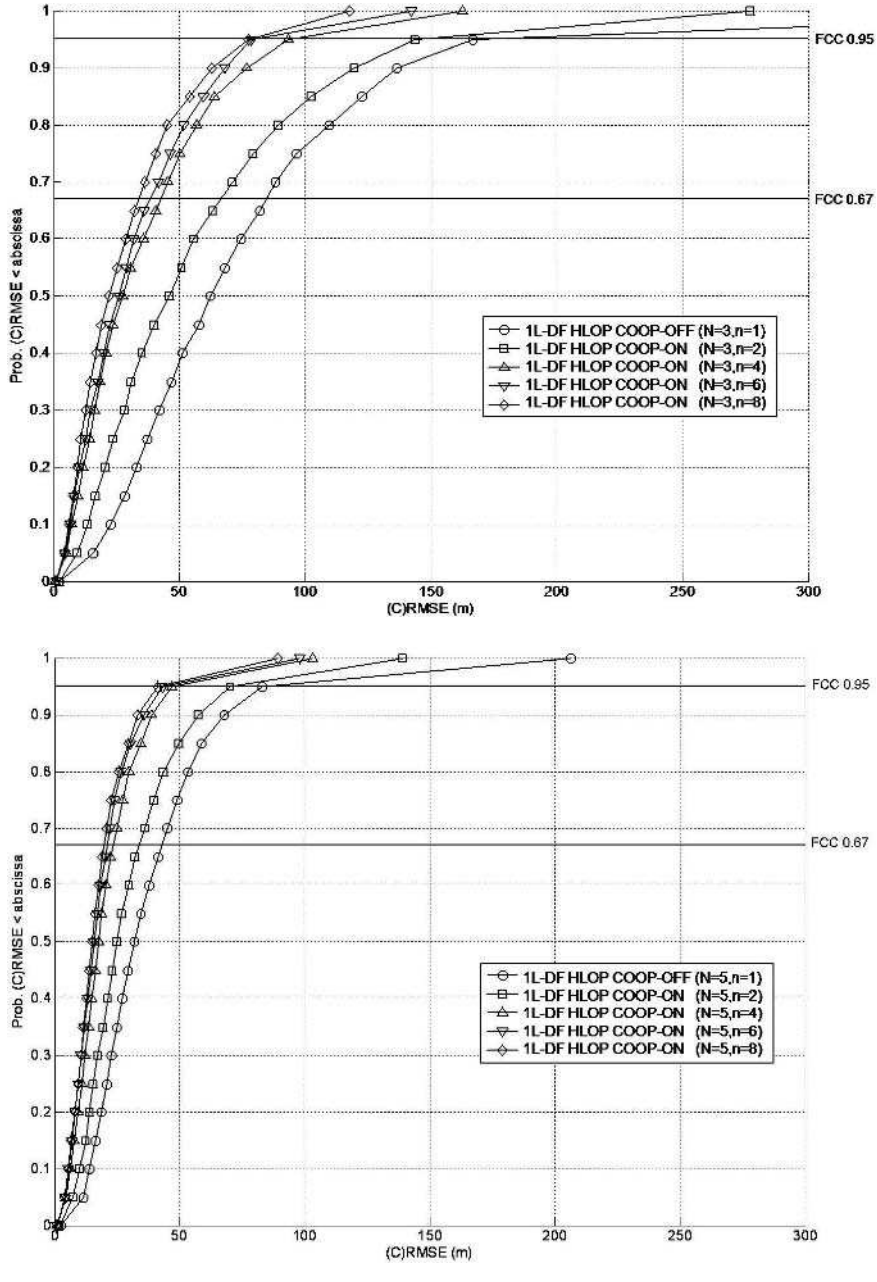


Figure 9.15 (C)RMSE vs. number of BSs (HLOP case). (S. Frattasi, M. Monti, "Ad-Coop positioning system (ACPS): Positioning for cooperative users in hybrid cellular ad-hoc networks", *European Transactions on Telecommunications Journal*, Wiley, vol. 19, no. 8, pp. 923–924, May, 2007. Reproduced in part by permission of © 2007 John Wiley & Sons Ltd.)

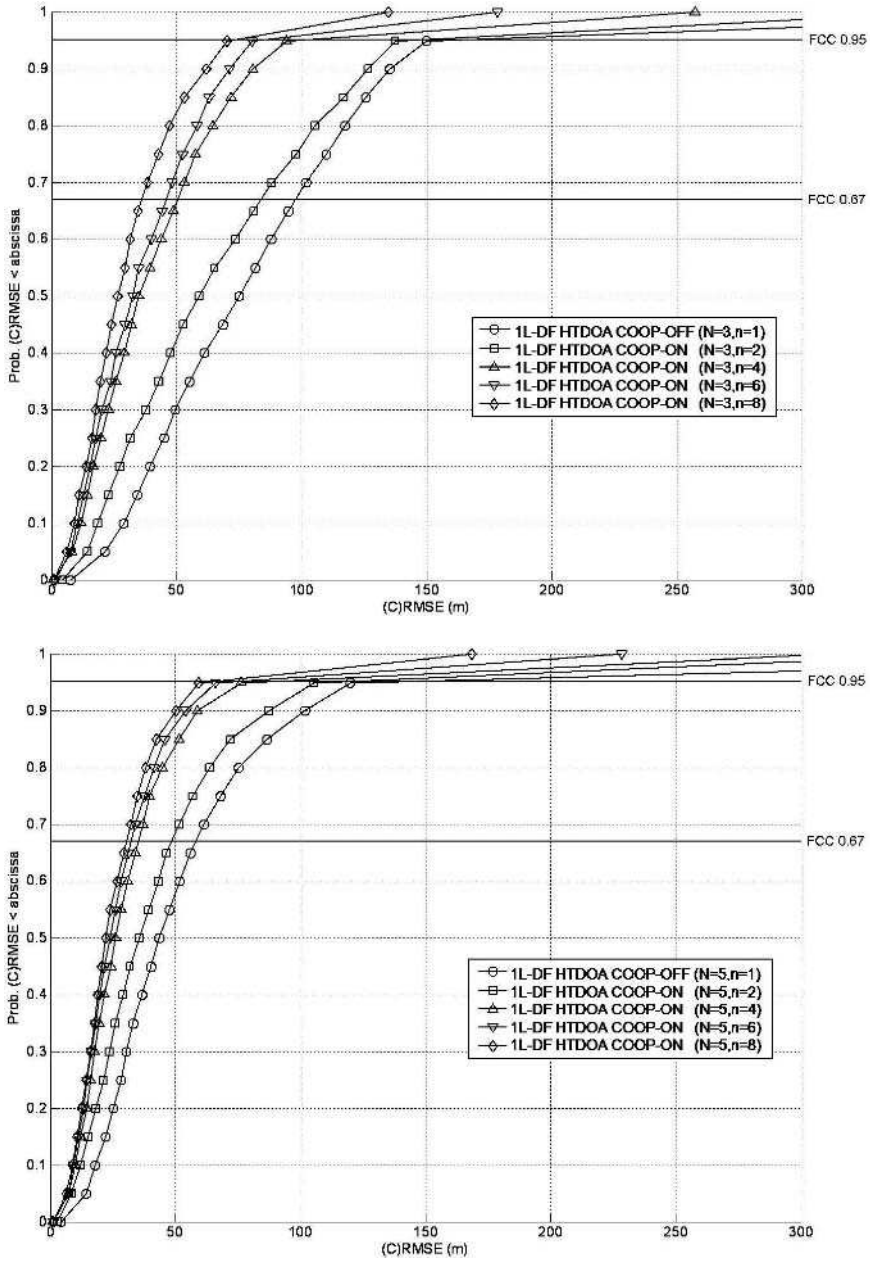


Figure 9.16 (C)RMSE vs. number of BSs (HTDOA case).

$[N = 5, n = 8]$, the location estimation accuracy is increased on average by about 13% (Table 9.3).

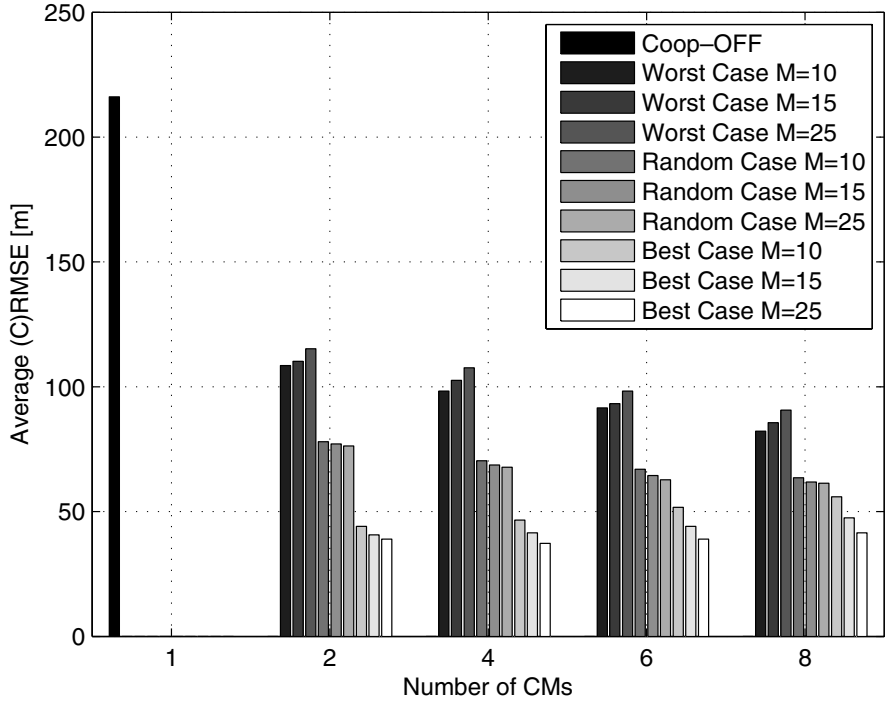


Figure 9.17 Average (C)RMSE vs. number of CMs and clustering method (HTAP case).

9.4.3.2.3 Dependence of performance on the clustering method Figure 9.17 shows a comparison, in terms of location estimation accuracy, between the results obtained with 1L-DF with and without cooperation for a variable number of CMs and different clustering methods, where the CH is placed at (500, 0) m ($P_{\text{los}} = 0$ for BS–CM links), the initial guesses are calculated by HTAP and the overall number of MSs in the surroundings of the CH (i.e., the candidate CMs) varies from 10 to 25. In particular, we compare the clustering method described in Section 9.4 (referred to as the "best case") with the case in which the CMs are a selection from the neighboring MSs which have been sorted in ascending order according to their SNR levels on the BS–MS links (referred to as the "worst case") and the case in which the choice of the CMs is made randomly (referred to as the "random case"). The following can be observed. (1) Regardless of the number of CMs involved and the clustering method adopted, the case with cooperation on always outperforms the case with cooperation off. In particular, the location estimation accuracy is increased on average by about 80% for the best case. (2) The best case always outperforms the other two cases; therefore, we can deduce that intelligent clustering of the CMs can lead to further location estimation enhancements. (3) The higher the number of candidate MSs, the higher the location estimation accuracy achieved for the best case. This is because

increasing the number of MSs in the surroundings of the CH offers a larger choice of candidate CMs, i.e., a larger range of SNR levels available. As expected, the reverse is valid for the worst case, while the results are practically invariant for the random case. (4) In contrast to the other two cases, the location estimation accuracy decreases slightly with the number of CMs for the best case. This is because the candidate CMs are sorted in decreasing order of SNR levels. Hence, there is a delicate balance when a new CM is chosen to be part of the cluster as the accuracy would tend to decrease owing to the decreasing SNR, although, at the same time, as explained in previous sections, the newly available measurements will create more constraints on the geometry of the problem. Nevertheless, in addition this minor trend, we can deduce that a few CMs can be enough to achieve the same accuracy as that obtained with many. This is a remarkable result, which implies that we can sensibly reduce the complexity of the localization system.

9.5 Conclusions

In this chapter, we have proposed an innovative solution for position determination in 4G wireless networks by introducing the cooperative mobile positioning approach. In particular, we have described a system, called COMET, which is a direct application of such an approach to cellular networks. The numerical results shown in the chapter have demonstrated that, regardless of the number of cooperating users or base stations available, COMET enhances the location estimation accuracy with respect to conventional hybrid positioning techniques in stand-alone cellular networks and, to a certain extent, it can achieve performance very much comparable to the GPS. Moreover, we have shown that an intelligent clustering of the users can further improve the accuracy of the localization system. In conclusion, this work has proved that the emerging paradigm of wireless cooperation has a beneficial impact on location estimation accuracy. Indirectly, this is also translated into an advantage from the point of view of handover, routing, beamforming and so on (see Chapter 3). In particular, within the framework of cognitive radio systems, such an advance could imply that spectrum holes could be detected more precisely.

References

3GPP Technical Specification Group (2005) *GSM/EDGE Radio Access Network: Radio Subsystem Link Control (3GPP Standard TS 05.08 V8.23.0)*, 3GPP.

3GPP Technical Specification Group (2006) *Technical Specification Group Radio Access Network. Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA): Release 7 (Draft Standard 3GPP TR 25.814 V7.1.0)*, 3GPP.

3GPP Technical Specification Group (2008a) *Radio Access Network: Stage 2 Functional Specification of User Equipment (UE) Positioning in UTRAN (3GPP Standard TS 25.305)*, 3GPP.

3GPP Technical Specification Group (2008b) *Services and System Aspects: Functional Stage 2 Description of Location Services (LCS) (3GPP Standard TS 23.271 V8.0.0)*, 3GPP.

3GPP Technical Specification Group (2008c) *Services and System Aspects: Location Services (LCS): Service Description: Stage 1 (3GPP Standard TS 22.071 V8.1.0)*, 3GPP.

3GPP Technical Specification Group (2008d) *Services and System Aspects: Universal Geographical Area Description (GAD) (3GPP Standard TS 23.032 V8.0.0)*, 3GPP.

3GPP Technical Specification Group (2009) *GSM/EDGE Radio Access Network: Radio Subsystem Synchronization (3GPP Standard TS 45.010 V8.4.0)*, 3GPP.

AAA (2009) American Automobile Association, `http://www.aaa.com` [10 March 2010].

Ad2Hand (2009) `http://www.ad2hand.com/` [10 March 2010].

Alamouti SM (1998) A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, **16**(8), 1451–1458.

Alhmiedat T and Yang S (2007) A survey: Localization and tracking mobile targets through wireless sensor networks. *Proceedings of the 8th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting*, Liverpool, UK.

Allen B, Dohler M, Okon E, Malik W, Brown A and Edwards D (2007) *Ultra-wideband Antennas and Propagation for Communications, Radar, and Imaging*, John Wiley & Sons, Ltd, Chichester.

Aloi D and Korniyenko O (2007) Comparative performance analysis of a Kalman filter and a modified double exponential filter for GPS-only position estimation of automotive platforms in an urban-canyon environment. *IEEE Transactions on Vehicular Technology*, **56**(5), part 2, 2880–2892.

Antonini M, Ruggieri M, Prasad R, Guida U and Corini G (2004) Vehicular remote tolling services using EGNOS. *Proceedings of the ION/IEEE Position Location and Navigation Symposium (PLANS 2004)*, Monterey, CA, pp. 375–379.

© 2010 John Wiley & Sons, Ltd

Arulampalam M, Maskell S, Gordon N, Clapp T, Sci D, Organ T and Adelaide S (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50**(2), 174–188.

Baert AE and Seme D (2004) Voronoi mobile cellular networks: Topological properties. *Proceedings of the International Symposium on Parallel and Distributed Computing in Association with HeteroPar'04*, Cork, Ireland, pp. 29–35.

Bai F, Sadagopan N and Helmy A (2003) IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM 2003*, San Francisco, CA, vol. 2, pp. 825–835.

Bandara U, Hasegawa M, Inoue M, Morikawa H and Aoyama T (2004) Design and implementation of a Bluetooth signal strength based location sensing system. *Proceedings of the Radio and Wireless Conference*, Atlanta, GA, pp. 319–322.

Bar-Shalom Y, Li X and Kirubarajan T (2001) *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, New York.

Bassiouni MA, Chiu MH, Loper M, Garnsey M and Williams J (1997) Performance and reliability analysis of relevance filtering for scalable interactive distributed simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **7**(3), 293–331.

Bluetooth Special Interest Group (2003) Bluetooth core specification v1.2.

Bolcskei H, Gesbert D and Paulraj AJ (2002) On the capacity of OFDM-based spatial multiplexing systems. *IEEE Transactions on Communications*, **50**(2), 225–234.

Borman S (2004) The Expectation Maximization Algorithm: A Short Tutorial. Lecture Notes, The Natural Language Group, University of South California

Bowditch N (2002) *American Practical Navigator: An Epitome of Navigation*, Paradise Cay Publications, pp. 99–104.

Box G and Jenkins G (1976) *Time Series Analysis, Forecasting and Control*, Holden-Day Series in Time Series Analysis, Holden-Day, San Francisco, CA.

Brenner P (1997) A technical tutorial on the IEEE 802.11 protocol, `http://sss-mag.com/pdf/802_11tut.pdf` [10 March 2010].

Brimicombe A (2002) GIS: Where are the frontiers now? *Proceedings of the 1st International Conference on GIS (GIS 2002)*, Manama, Bahrain, pp. 33–45.

Butler J (2003) Mobile robots as gateways into wireless sensor networks, `http://www.linuxdevices.com/articles/AT2705574735.htm` [10 March 2010].

Caffery JJ (1999) *Wireless Location in CDMA Cellular Radio Systems*, Kluwer, Norwell, MA, Chapter 7.

Caffery JJ and Stüber GL (1998) Overview of radiolocation in CDMA cellular systems. *IEEE Communications Magazine*, **36**(4), 38–45.

Caffery JJ and Stüber GL (1998) Subscriber location in CDMA cellular networks. *IEEE Transactions on Vehicular Technology*, **47**(2), 406–416.

Capkun S, Hamdi M and Hubaux JP (2002) GPS-free positioning in mobile ad-hoc networks. *Cluster Computing*, **5**(2), 157–167.

Casas R, Marco A, Guerrero JJ and Falco J (2006) Robust estimator for non-line-of-sight error mitigation in indoor localization. *EURASIP Journal of Applied Signal Processing*, **2006**(1), 1–8.

CCIR (1989) *Spectrum Conserving Terrestrial Frequency Assignments for Given Frequency Distance Separations*, CCIR.

CELLO (2003) Cellular network optimization based on mobile location, `http://www.telecom.ntua.gr/cello` [17 August 2009].

Celusion (2009) FDMA, TDMA, CDMA & GSM – What is the future?, `http://www.celusion.com/whitepapers/fdma_tdma_cdma.htm` [10 March 2010].

Chan YT and Ho KC (1994) A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, **42**(8), 1905–1915.

Chan YT, Hang HYC and Ching PC (2006a) Exact and approximate maximum likelihood localization algorithms. *IEEE Transactions on Vehicular Technology*, **55**(1), 10–16.

Chan YT, Tsui WY, So HC and Ching PC (2006b) Time of arrival based localization under NLOS conditions. *IEEE Transactions on Vehicular Technology*, **55**(1), 17–24.

Chaudhari N (2007) Location-aided operation of future mobile wireless networks, White Paper, Wipro Technologies.

Chen PC (1999) A non-line-of-sight error mitigation algorithm in location estimation. *Proceedings of the IEEE International Conference on Wireless Communications and Networking (WCNC)*, New Orleans, LA, vol. 1, pp. 316–320.

Chen C and Feng K (2005a) Hybrid location estimation and tracking system for mobile devices. *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC-Spring)*, Stockholm, Sweden, vol. 4, pp. 2648–2652.

Chen CL and Feng KT (2005b) An efficient geometry-constrained location estimation algorithm for NLOS environments. *Proceedings of the IEEE International Conference on Wireless Networks, Communication, Mobile Computing*, Hawaii, pp. 244–249.

Cheng C and Sahai A (2004) Estimation bounds for localization. *Proceedings of the IEEE International Conference on Sensors and Ad-Hoc Communications and Networks (SECON)*, Santa Clara, CA, pp. 415–424.

Cheng YC, Chawathe Y and Krumm J (2005) Accuracy characterization for metropolitan-scale Wi-Fi localization. Technical Report IRS-TR-05-003, Intel Research.

Chernick MR (1999) *Bootstrap Methods: A Practitioner's Guide*, John Wiley & Sons, New York.

Chiang CC (1998) Wireless network multicasting. Ph.D. thesis, University of California, Los Angeles.

Chiu MH and Bassiouni M (2000) Predictive schemes for handoff prioritization in cellular networks based on mobile positioning. *IEEE Journal on Selected Areas in Communications*, **18**(3), 510–522.

Choi S and Shin G (1998) Predictive and adaptive bandwidth reservation for hand-offs in QoS-sensitive cellular networks. *ACM SIGCOMM*, Vancouver, Canada, vol. 28, issue 4, pp. 155–166.

Chon HD, Jun S, Jung H and An SW (2004) Using RFID for accurate positioning. *Journal of Global Positioning Systems*, **3**(1–2), 32–39.

Cohn S, Sivakumaran N and Todling R (1994) A fixed-lag Kalman smoother for retrospective data assimilation. *Monthly Weather Review*, **122**(12), 2838–2867.

Coleman TF and Li Y (1996) An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, **6**, 418–445.

Cong L and Zhuang W (2001) Non-line-of-sight error mitigation in TDOA mobile location. *Proceedings of the IEEE Global Telecommunication Conference (GLOBECOM)*, San Antonio, TX, vol. 1, pp. 680–684.

Cong L and Zhuang W (2002) Hybrid TDOA/AOA mobile user location for wideband CDMA cellular systems. *IEEE Transactions on Wireless Communications*, **1**, 439–447.

Cong L and Zhuang W (2004) Non-line-of-sight error mitigation in mobile location. *Proceedings of the 23rd IEEE Annual Joint Conference of the Computer and Communications Societies (INFOCOM)*, Hong Kong, pp. 650–659.

Cong L and Zhuang W (2005) Nonline-of-sight error mitigation in mobile location. *IEEE Transactions on Wireless Communications*, **4**(2), 560–573.

David L (2009) David L. Wilson's GPS accuracy Web page, `http://users.erols.com/dlwilson/gps.htm` [10 March 2010].

Deffenbaugh M, Bellingham J and Schmidt H (1996) The relationship between spherical and hyperbolic positioning. *Conference Proceedings OCEANS '96, MTS/IEEE, Prospects for the 21st Century*, Fort Lauderdale, FL, vol. 2, pp. 590–595.

Deng P and Fan P (2000) An AOA assisted TOA positioning system. *Proceedings of the World Computer Congress/International Conference on Communication Technology (WCC/ICCT)*, Beijing, vol. 3, pp. 1501–1504.

Denis B, Pierrot JB and Abou-Rjeily C (2006) Joint distributed synchronization and positioning in UWB ad hoc networks using TOA. *IEEE Transactions on Microwave Theory and Techniques*, **54**(4), 1896–1911.

Dennis JE and Schnabel RB (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-hall, Englewood Cliffs, NJ.

Departments of Defense and of Transportation (2001) *2001 Federal Radionavigation Systems*. Department of Defense and Department of Transportation, National Technical Information Service, Springfield, VA.

Dizdarevic V and Witrisal K (2006) On impact of topology and cost function on LSE position determination in wireless networks. *Proceedings of the Workshop on Positioning, Navigation, and Communication (WPNC)*, Hannover, Germany, pp. 129–138.

Dodgeball (2009) `http://www.dodgeball.com` [13 April 2009].

Doucet A, Freitas N de and Gordon N (2001) *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.

Drane C, Macnaughtan M and Scott C (1998) Positioning GSM telephones. *IEEE Communications Magazine*, **36**(4), 46–59.

Eltaher A (2009) *Electro-Optical Ranging for Short Range Applications: Design and Realization Aspects*. Shaker.

Erceg V (2001) Channel model for fixed wireless application. Technical report, IEEE 802.16 Working group.

Erceg V, Greenstein LJ, Tjandra SY, Parkoff SR, Gupta A, Kulic B, Julius AA and Bianchi R (1999a) An empirically based path loss model for wireless channels in suburban environments. *IEEE Journal on Selected Areas in Communications*, **17**(7), 1205–1211.

Erceg V, Michelson DG, Ghassemzadeh SS, Greenstein LJ, Rustako AJ, Guerlain PB, Dennison MK, Roman RS, Barnickel DJ, Wang SC and Miller RR (1999b) A model for the multipath delay profile of fixed wireless channels. *IEEE Journal on Selected Areas in Communications*, **17**(3), 399–410.

ETSI (1998) Universal Mobile Telecommunications System (UMTS): Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.1.0). Technical report.

EUROPA (2003) Commission pushes for rapid deployment of location enhanced 112 emergency services. DN: IP/03/1122.

European Commission Information and Media (2005) eSafety: Improving road safety using information & communication technologies. Fact Sheet 48, `http://www.esafetysupport.org/download/European_Commission/048-esafety.pdf` [10 March 2010].

Farrell J and Barth M (1999) *The Global Positioning System and Inertial Navigation*, McGraw-Hill.

FCC (1999) FCC acts to promote competition and public safety in enhanced wireless 911 services. DC: WT Rep. 99-27.

FCC (2001) FCC wireless 911 requirements. Fact sheet WTB/Policy.

Feess WA and Stephens SG (1987) Evaluation of GPS ionospheric time-delay model. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-23**, 332–338.

Feldmann S, Kyamakya K, Zapater A and Lue Z (2003) An indoor Bluetooth-based positioning system: Concept, implementation and experimental evaluation. *Proceedings of the International Conference on Wireless Networks*, Las Vegas, NV, pp. 109–113.

Figueiras J (2008) Accuracy enhancements for positioning of mobile devices in wireless communication networks. Ph.D. thesis, Aalborg University.

Figueiras J and Schwefel HP (2008) Wireless positioning based on a segment-wise linear approach for modeling the target trajectory. *IEEE Global Communications Conference (IEEE GLOBECOM 2008)*, New Orleans, LA, pp. 4825–4829.

Figueiras J, Schwefel HP and Kovacs I (2005) Accuracy and timing aspects of location information based on signal-strength measurements in Bluetooth. *16th International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC 2005)*, Berlin, Germany, vol. 4, pp. 2685–2690.

Foschini GJ (1996) Layered space–time architecture for wireless communication in a fading environment when using multiple antennas. *Bell Labs Technical Journal*, **6**, 41–59.

Foschini GJ and Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, **6**, 311–335.

Fox D, Hightower J, Liao L, Schulz D and Borriello G (2003) Bayesian filtering for location estimation. *IEEE Pervasive Computing*, **2**(3), 24–33.

Frattasi S (2007) Link layer techniques enabling cooperation in fourth generation (4G) wireless networks. Ph.D. thesis, Aalborg University.

Frattasi S and Figueiras J (2007) *Ad-Coop Positioning System: Using an Embedded Kalman Filter Data Fusion*, CRC Press, pp. 120–131.

Frattasi S and Monti M (2006) On the use of cooperation to enhance the location estimation accuracy. *Proceedings of the IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Valencia, Spain, pp. 545–549.

Frattasi S and Monti M (2007a) Ad-coop positioning system (ACPS): Positioning for cooperating users in hybrid cellular ad-hoc networks. *Wiley European Transactions on Telecommunications*, **19**(8), 929–934.

Frattasi S and Monti M (2007b) Cooperative mobile positioning in 4G wireless networks, *Cognitive Wireless Networks*, Springer, pp. 213–233.

Frattasi S, Monti M and Prasad R (2006) A cooperative localization scheme for 4G wireless communications. *Proceedings of the IEEE Radio & Wireless Symposium (RWS)*, San Diego, CA, pp. 287–290.

Gezici S (2008) A survey on wireless position estimation. *Springer Wireless Personal Communications*, **44**(3), 263–282.

Gezici S and Sahinoglu Z (2004) UWB geolocation techniques for IEEE 802.15.4a personal area networks. MERL technical report.

Goldsmith A (2005) *Wireless Communications*. Cambridge University Press.

Gonzalez-Castano F and Garcia-Reinoso J (2002) Bluetooth location networks. *IEEE Global Telecommunications Conference, GLOBECOM 2002*, Taipei, Taiwan, vol. 1, pp. 233–237.

Google (2006) Review guide. Google Maps for Mobile (beta), `http://maps.google.com` [10 March 2010].

Google (2007) Google Maps for Mobile with My Location (beta). Online video, `http://www.youtube.com/watch?v=v6gqipmbcok` [10 March 2010].

Gordon N, Salmond D and Smith A (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, **140**(2), 107–113.

Graphisoft R&D (2007) Location-Based Construction Management., `http://martasari.files.wordpress.com/2007/09/c2007_print_a4.pdf` [10 March 2010].

Green J, Betti D and Davison J (2000) *Mobile location services: Market strategies*. Technical Report, Ovum.

Greenstein L, Yeh VEY and Clark M (1997) A new path-gain/delay-spread propagation model for digital cellular channels. *IEEE Transactions on Vehicular Technology*, **46**(2), 477–485.

Grewal M and Andrews A (2001) *Kalman Filter: Theory and Practice Using MATLAB*, 2nd edn., John Wiley & Sons.

Grewal MS, Weill LR and Andrews AP (2007) *Global Positioning Systems, Inertial Navigation, and Integration*, 2nd edn., John Wiley & Sons.

Grubbs F (1969) Procedures for detecting outlying observations in samples. *Technometrics*, **11**(1), 1–21.

Gustafsson F and Gunnarsson F (2005a) Mobile positioning using wireless networks: Possibilites and fundamental limitations based on available wireless network measurements. *IEEE Signal Processing Magazine*, **22**(4), 41–53.

Gustafsson F and Gunnarsson F (2005b) Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Processing Magazine*, **22**, 41–53.

Guvenc I and Chong CC (2009) A survey on TOA based wireless localization and NLOS mitigation techniques. *IEEE Communications Surveys and Tutorials*, **11**(3), 107–124.

Guvenc I, Chong CC and Watanabe F (2007) NLOS identification and mitigation for UWB localization systems. *Proceedings of the IEEE International Conference on Wireless Communication and Networking (WCNC)*, Hong Kong, pp. 1571–1576.

Guvenc I, Chong CC, Watanabe F and Inamura H (2008) NLOS identification and weighted least squares localization for UWB systems using multipath channel statistics. *EURASIP Journal of Advances in Signal Processing (Special Issue on Signal Processing for Location Estimation and Tracking in Wireless Environments)*, **2008**(1), 1–14.

Hasu V (2007) Radio resource management in wireless communication: Beamforming, transmission power control, and rate allocation. Ph.D. thesis, Helsinki University of Technology.

Heath R and Paulraj A (2000) Switching between multiplexing and diversity based on constellation distance. *Allerton Conference on Communication, Control and Computing*, Morticello, IL, pp. 212–219.

Heidari G (2008) *WiMedia UWB for W-USB and Bluetooth Interpretation of Standards, Regulations and Applications*, John Wiley & Sons.

Heinzelman W, Chandrakasan A and Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, **1**(4), 660–670.

Hightower J, Brumitt B and Borriello G (2002) The location stack: A layered model for location in ubiquitous computing. *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, Callicoon, NY, pp. 22–28.

Hong D and Rappaport SS (1986) Traffic model and performance analysis for cellular mobile radio telephone system with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology*, **35**, 77–92.

Howard A, Matarić M and Sukhatme G (2003) Cooperative relative localization for mobile robot teams: An ego-centric approach. *Proceedings of the Naval Research Laboratory Workshop on Multi-Robot Systems*, Washington, DC, pp. 65–76.

Huber PJ (1964) Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**(1), 73–101.

IEEE (2002) 802 IEEE standard for local and metropolitan area networks: Overview and architecture. Technical report, IEEE.

IEEE (2005) IEEE 802.15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs).

IF (2009) Issued and pending patents offered to Microsoft Corporation, `http://innfundllc.com/press/pr_10.html` [7 July 2009].

Jacob K (2007) Location-based services, `http://www.kenneyjacob.com/2007/05/13/location-based-services` [10 March 2010].

Jakes WC (1994) *Microwave Mobile Communications*, John Wiley & Sons.

Jiao W, Jiang P, Liu R, Wang W and Ma Y (2008) Providing location service for mobile WiMAX. *IEEE International Conference on Communications, 2008 (ICC '08)*, Beijing, China, pp. 2685–2689.

Jones Q and Grandhi S (2005) P3 systems: Putting the place back into social networks. *IEEE Internet Computing*, **9**(5), 38–46.

Jourdan DB and Roy N (2006) Optimal sensor placement for agent localization. *Proceedings of the IEEE Position, Location, and Navigation Symposium (PLANS)*, San Diego, CA, pp. 128–139.

Jourdan DB, Dardari D and Win MZ (2006) Position error bound for UWB localization in dense cluttered environments. *Proceedings of the IEEE International Conference on Communications (ICC)*, Istanbul, vol. 8, pp. 3705–3710.

Julier S and Uhlmann J (1997) A new extension of the Kalman filter to nonlinear systems. *International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, FL.

Jung P (1997) *Analyse und Entwurf digitaler Mobilfunksysteme*. Teubner.

Kailath T, Sayed AH and Hassibi B (2000) *Linear Estimation*, Prentice Hall Information and Systems Science Series, Prentice Hall, Upper Saddle River, NJ.

Kalman R (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**(1), 35–45.

Kaplan E and Hegarty C (2006) *Understanding GPS: Principles and Application*, Artech House.

Kim W, Lee JG and Jee GI (2006) The interior-point method for an optimal treatment of bias in trilateration location. *IEEE Transactions on Vehicular Technology*, **55**(4), 1291–1301.

Kim S, Jeon H, Lee H and Ma JS (2008) Robust transmission power and position estimation in cognitive radio. *Springer Lecture Notes in Computer Science*, **5200**, 719–728.

kingpin (1998) MAC address cloning, `http://www.packetstormsecurity.org/docs/hack/mac_address_cloning.pdf` [10 March 2010].

Kivanc D and Liu H (2000) Subcarrier allocation and power control for OFDMA. *IEEE Signals, Systems and Computers*, Pacific Grove, CA, pp. 147–151.

Kivera (2002) Value & applications for location-based services. An overview of opportunities & challenges for deploying state-of-the-art location technologies, `http://www.kivera.com/downloads/whitepaper.pdf` [10 March 2010].

Klobuchar J (1976) Ionospheric time-delay corrections for advanced satellite ranging systems. *NATO AGARD Conference Proceedings*, Paris, France, vol. 209.

Knapp C and Carter G (1976) The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **24**(4), 320–327.

Knight NL and Wang J (2009) A comparison of outlier detection procedures and robust estimation methods in GPS positioning. *Journal of Navigation, Royal Institute of Navigation*, **62**(4), 699–709.

Ko Y and Vaidya N (1998) Location aided routing (LAR) in mobile ad-hoc networks. *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Dalla, TX, pp. 66–75.

Koeppel I (2002) What are location services? – from a GIS perspective, `http://lbs360.directionsmag.com/LBSArticles/ESRI.What%20are%20LS%20Whitepaper.pdf` [10 March 2010]

Küpper A (2005) *Location-Based Services: Fundamentals and Operation*, John Wiley & Sons, Ltd, Chichester.

Kyammaka K and Jobmann K (2005) Location management in cellular networks: Classification of the most important paradigms, realistic simulation framework, and relative performance analysis. *IEEE Transactions on Vehicular Technology*, **54**, 687–708.

Larsson EG (2004) Cramer–Rao bound analysis of distributed positioning in sensor networks. *IEEE Signal Processing Letters*, **11**(3), 334–337.

Lathi MJ (2000) IEEE 802.11 wireless LAN. *Proceedings of the Seminar on Internetworking on Ad-Hoc Networking*, Helsinki University of Technology.

LaViola J (2003) Double exponential smoothing: An alternative to Kalman filter-based predictive tracking. *Proceedings of the Immersive Projection Technology and Virtual Environments 2003*, pp. 199–206.

Lawton MC, Davies RL and McGeehan JP (1991) An analytical model for indoor multipath propagation in the picocellular environment. *Proceedings of the 6th International Conference on Mobile Radio and Personal Communications*, Coventry, UK, pp. 1–8.

Levine DA, Akyildiz IF and Naghshineh M (1997) A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept. *IEEE/ACM Transactions on Networking*, **5**, 1–12.

Li X (2006) An iterative NLOS mitigation algorithm for location estimation in sensor networks. *Proceedings of the IST Mobile and Wireless Communications Summit*, Mykonos, Greece.

Li Z, Trappe W, Zhang Y and Nath B (2005) Robust statistical methods for securing wireless localization in sensor networks. *Proceedings of the IEEE International Symposium on Information Processing in Sensor Networks (IPSN)*, Los Angeles, CA, pp. 91–98.

Liberti J and Rappaport T (1992) Statistics of shadowing in indoor radio channels at 900 and 1900 MHz. *IEEE Military Communications Conference, 1992, MILCOM '92, Conference Record, Communications – Fusing Command, Control and Intelligence*, vol. 3, pp. 1066–1070.

Lima P (2007) A Bayesian approach to sensor fusion in autonomous sensor and robot networks. *IEEE Instrumentation and Measurement Magazine*, **10**(3), 22–27.

Liu H, Darabi H, Banerjee P and Liu J (2007) Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, **37**(6), 1067–1080.

Lu S and Bharghavan V (1996) Adaptive resource management algorithms for indoor mobile computing environment. *ACM SIGCOMM*, Stanford, CA, pp. 231–242.

Luba O, Boyd L, Gower A and Crum J (2005) GPS III system operations concepts. *IEEE Aerospace and Electronic Systems Magazine*, **20**, 10–18.

Madsen K, Nielsen H and Tingleff O (2004) *Methods for Non-linear Least Squares Problems*, 2nd edn, Technical University of Denmark, Lyngby.

Malm A (2006) GPS and Galileo in Mobile Handsets. *LBS Research Series*, Technical Report, Berg Insight.

Mao G, Fidan B and Anderson B (2007) Wireless sensor network localization techniques. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, **51**(10), 2529–2553.

MapInfo (2009) `http://www.pbinsight.com` [10 March 2010].

Maxim (2003) An introduction to spread-spectrum communications, `http://www.maxim-ic.com/appnotes.cfm/an_pk/1890/` [10 March 2010].

Meel J (1999) Spread spectrum – introduction, De Nayer Instituut, `http://sss-mag.com/pdf/Ss_jme_denayer_intro_print.pdf` [10 March 2010].

Merwe R and Rebel E (2006) Recursive Bayesian estimation library, `http:\\choosh.csee.ogi.edu/rebel` [10 March 2010].

Mhatre V and Rosenberg C (2004) Design guidelines for wireless sensor networks: Communication, clustering and aggregation. *Elsevier Science Ad-Hoc Networks Journal*, **2**(1), 45–63.

Mhatre V and Rosenberg C (2005) Energy and cost optimizations in wireless sensor networks: a survey. *Performance Evaluation and Planning Methods for the Next Generation Internet* (eds Girard A, Sansò B and Vázquez-Abad F), Springer, pp. 227–248.

Miao H (2007) Channel estimation and positioning for multiple antenna systems. Ph.D. thesis, University of Oulu.

Mitola J and Maguire GQ (1999) Cognitive radio: Making software radios more personal. *IEEE Personal Communications*, **6**(4), 13–18.

MobileIN (2009) Value-added services, `http://www.mobilein.com/what_is_a_VAS.htm` [10 March 2010].

MobileRobots (2009) `http://www.mobilerobots.com/` [10 March 2010].

Mohay G, Anderson A, Collie B, De Vel O and McKemmish R (2003) *Computer and Intrusion Forensics*, Artech House.

Molisch AF (2005) *Wireless Communications*, John Wiley & Sons.

Montalvo UD, Ballon P and de Kar EV (2003) Business models for location-based services. *Proceedings of the 6th AGILE Conference on GIScience*, Lyon, France.

Mourikis A and Roumeliotis S (2006a) Optimal sensor scheduling for resource-constrained localization of mobile robot formations. *IEEE Transactions on Robotics*, **22**(5), 917–931.

Mourikis A and Roumeliotis S (2006b) Performance analysis of multirobot cooperative localization. *IEEE Transactions on Robotics*, **22**(4), 666–681.

Muhammad W, Grosicki E, Abed-Meraim K, Delmas JP and Desbouvries F (2002) Uplink versus downlink wireless mobile positioning in UMTS cellular radio systems. *Proceedings of the European Signal Processing Conference (EUSIPCO'02)*, Toulouse, France, vol. 2, pp. 373–376.

Murphy K (1988) Switching Kalman filters. Technical report, Department of Computer Science, University of California, Berkeley.

Musolesi M, Hailes S and Mascolo C (2004) An ad hoc mobility model founded on social network theory. *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems 2004, MSWiM'04*, Venice, Italy, pp. 20–24.

Nepper P, Treu G and Küpper A (2008) Adding speech to location-based services. *Springer Wireless Personal Communications, Special Issue, Towards Global & Seamless Personal Navigation*, **44**(3), 245–261.

Niedzwiadek H (2002) Where's the value in location services, `http://lbs360.directionsmag.com/LBSArticles/HN.Where's%20the%20Value%20in%20Location%20Services.pdf` [10 March 2010].

NIST/SEMATECH (2007) *e-Handbook of Statistical Methods*, NIST, `http://www.itl.nist.gov/div898/handbook/` [10 March 2010].

Nosratinia A, Hunter TE and Hedayat A (2004) Cooperative communication in wireless networks. *IEEE Communications Magazine*, **42**(10), 74–80.

Obayashi S and Zander J (1998) A body-shadowing model for indoor radio communication environments. *IEEE Transactions on Antennas and Propagation*, **46**(6), 920–927.

OGC (2003) Open GIS location services (OPENLS): Core services. OGC 03-006r1.

Pantel L and Wolf LC (2002) On the suitability of dead reckoning schemes for games. *Proceedings of the 1st workshop on Network and system support for games (NetGames)*, Braunschweig, Germany, pp. 79–84.

Parker L (1996) Multi-robot team design for real-world applications. *Distributed Autonomous Systems 2*, Springer-Verlag, Tokyo, pp. 91–102.

Parkinson B and Spilker J (1996) *The Global Positioning System: Theory and Applications*, American Institute of Aeronautics and Astronautics.

Parsons J (2000) *The Mobile Radio Propagation Channel*, John Wiley & Sons.

Patwari N, Ash J, Kyperountas S, Hero A, Moses R and Correal N (2005) Locating the nodes: Cooperative localization in wireless sensor networks. *IEEE Signal Processing Magazine*, **22**(4), 54–69.

Paulraj A and Kailath T (1994) Increasing capacity in wireless broadcast systems using distributed transmission/directional reception. US Patent 5345599.

Paulraj A, Nabar R and Gore D (2003) *Introduction to Space–Time Wireless Communications*, Cambridge University Press.

Pels M, Barhorst J, Michels M, Hobo R and Barendse J (2005) Tracking people using Bluetooth: Implications of enabling Bluetooth discoverable mode. Technical Report, University of Amsterdam.

Penrose R (1955) A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, **51**, 406–413.

Peressini AL, Sullivan FE and Uhl JJ (1988) *The Mathematics of Nonlinear Programming*, Springer.

Petrus P (1999) Robust Huber adaptive filter. *IEEE Transactions on Signal Processing*, **47**(4), 1129–1133.

Prasad R, Mohr W and Konhauser W (2000) *Third Generation Mobile Communication Systems*, Artech House.

Prasad R, Rahman MI, Das SS and Marchetti N (2009) *Single- and Multi-Carrier MIMO Transmission for Broadband Wireless Systems*, River.

Proakis JG (1995) *Digital Communications*, Prentice Hall.

Qi Y, Kobayashi H and Suda H (2006) Analysis of wireless geolocation in a non-line-of-sight environment. *IEEE Transactions on Wireless Communication*, **5**(3), 672–681.

Rappaport TS (1996) *Wireless Communications Principles and Practice*, Prentice Hall.

Rhee W and Cioffi JM (2000) Increase in capacity of multiuser OFDM system using dynamic subchannel allocation. *IEEE Vehicular Technology Conference Spring'00*, Tokyo, pp. 1085–1089.

Riba J and Urruela A (2004) A non-line-of-sight mitigation technique based on ML-detection. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Quebec, Canada, vol. 2, pp. 153–156.

Rios S (2009) Location based services: Interfacing to a mobile positioning center, `http://www.wirelessdevnet.com/channels/lbs/features/lbsinterfacing.html` [10 March 2010].

Ripley BD (2004) Robust statistics. `http://www.stats.ox.ac.uk/pub/StatMeth` [10 March 2010].

Rosa FD (2007) Cooperative mobile positioning and tracking in hybrid WIMAX/WLAN networks. Master's thesis, Aalborg University.

Rosa FD, Simone G, Laurent P, Charaffedine R, Mahmood N, Pietrarca B, Kyritsi P, Marchetti N, Perrucci G, Figueiras J and Frattasi S (2007a) Emerging directions in wireless location: Vista from the COMET project. *Proceedings of the 16th IST Mobile & Wireless Communications Summit*, Budapest, Hungary, pp. 1–5.

Rosa FD, Wardana S, Mayorga C, Naima GSM, Figueiras J and Frattasi S (2007b) Experimental activity on cooperative mobile positioning in indoor environments. *Proceedings of the 2nd IEEE Workshop on Advanced Experimental Activities on Wireless Networks & Systems (EXPONWIRELESS)*, Helsinki, Finland, pp. 1–5.

Sari H and Karam G (1996) Orthogonal frequency-division multiple access for the return channel on CATV networks. *Proceedings of the International Conference on Telecommunications (IEEE ICT)*, Istanbul, Turkey, pp. 602–607.

Sayed A and Yousef NR (2003) Wireless location, *Wiley Encyclopedia of Telecommunications* (ed. Proakis J), John Wiley & Sons, New York.

Sayed A, Tarighat A and Khajehnouri N (2005a) Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Processing Magazine*, **22**(4), 24–40.

Sayed AH, Tarighat A and Khajehnouri N (2005b) Network-based wireless location. *IEEE Signal Processing Magazine*, **22**(4), 24–40.

Schiller J and Voisard A (2004) *Location-Based Services*, Morgan Kaufmann.

SDMA Toolbox for IT (2009) `http://it.toolbox.com/wiki/index.php/SDMA` [10 March 2010].

Second Life (2009) `http://secondlife.com` [10 March 2010].

Senarath G (2006) Multi-hop relay system evaluation methodology (channel model and performance metric), IEEE 802.16J-06/013R3.

Shiode N, Li C, Batty M, Logley P and Maguire D (2004) The impact and penetration of location-based services. *CRC Press Telegeoinformatics: Location-Based Computing and Services*, pp. 349–366.

Skyhook wireless (2009) `http://www.skyhookwireless.com` [10 March 2010].

Soork AS, Saadat R and Tadaion AA (2008) Cooperative mobile positioning based on received signal strength. *International Symposium on Telecommunications, 2008 (IST 2008)*, Tehran, Iran, pp. 273–277.

Steinfield C (2004) *The Development of Location Based Services in Mobile Commerce*. Springer, pp. 177–197.

Steininger S, Neun M and Edwardes A (2006) Foundations of location-based services. Technical Report, University of Zurich.

Sun GL and Guo W (2004) Bootstrapping M-estimators for reducing errors due to non-line-of-sight (NLOS) propagation. *IEEE Communications Letters*, **8**(8), 509–510.

Talukdar AK, Badrinath BR and Acharya A (1997) On accommodating mobile hosts in an integrated services packet network. *Sixteenth Annual Joint Conference of the IEEE Computer and Communication Societies (IEEE INFOCOM'97)*, Kobe, Japan, pp. 1048–1055.

Tarokh V, Seshadri N and Calderbank AR (1998) Space–time codes for high data rate wireless communications: Performance criterion and code construction. *IEEE Transactions on Information Theory*, **44**, 744–765.

Taylor J (2002) Robust modelling using the t-distribution. *Proceedings of the Australasian Genstat Conference*, Busselton, Australia.

Tejavanija K, Williamson K and Kang J (2003) Development of location-based facilities management system for mobile devices. *Proceedings of the 1st China–U.S. Relations Conference*, College Station, TX.

Telatar I (1999) Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, **10**(6), 585–595.

Telenity (2009) `http://www.telenity.com` [10 March 2010].

Todling R, Cohn S and Sivakumaran N (1997) Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Monthly Weather Review*, **126**(8), 2274–2286.

Tripathi ND, Reed JH and VanLandingham HF (1998) Handoff in cellular systems. *IEEE Personal Communications*, **5**(6), pp. 26–37.

TruePosition (2009) `http://www.trueposition.com` [10 March 2010].

Van de Kar EAM (2004) Designing mobile information services: An approach for organisations in a value network. Ph.D. thesis, Delft University of Technology.

Vanderveen M, Papadias C and Paulraj A (1997) Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array. *Communications Letters, IEEE*, **1**(1), 12–14.

Venkatesh S and Buehrer RM (2006) A linear programming approach to NLOS error mitigation in sensor networks. *Proceedings of the IEEE International Symposium on Information Processing in Sensor Networks (IPSN)*, Nashville, TN, pp. 301–308.

Venkatesh S and Buehrer RM (2007) NLOS mitigation using linear programming in ultrawideband location-aware networks. *IEEE Transactions on Vehicular Technology*, **56**(5), 3182–3198.

Venkatraman S and Caffery J (2004) Hybrid TOA/AOA techniques for mobile location in non-line-of-sight environments. *Proceedings of the Wireless Communications and Networking Conference (WCNC)*, Atlanta, GA, vol. 1, pp. 274–278.

Virrantaus K, Veijalainen J, Markkula J, Katanosov A, Garmash A, Tirri H and Terziyan V (2001) Developing GIS-supported location-based services. *Proceedings of the 1st International Workshop on Web Geographical Information Systems (WGIS 2001)*, Kyoto, Japan, vol. 2, pp. 66–75.

Vossiek M, Wiebking L, Gulden P, Wieghardt J, Hoffmann C and Heide P (2003) Wireless local positioning. *IEEE Microwave Magazine*, **4**, 77–86.

Waadt A, Hessamian-Alinejad A, Wang S, Statnikov K, Bruck G and Jung P (2008) Mobile-assisted positioning in GSM networks. *Proceedings of the International Workshop on Signal Processing and its Applications (WoSPA2008)*, Sharjah, UAE.

Wang X, Wang Z and Dea BO (2003) A TOA based location algorithm reducing the errors due to non-line-of-sight (NLOS) propagation. *IEEE Transactions on Vehicular Technology*, **52**(1), 112–116.

Wang Z, Tameh E and Nix A (2004) Statistical peer-to-peer channel models for outdoor urban environments at 2 GHz and 5 GHz. *Proceedings of the 60th IEEE Vehicular Technology Conference (VTC-Fall)*, Los Angeles, CA, vol. 7, pp. 5101–5105.

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

Wei D and Chan A (2005) A survey on cluster schemes in ad-hoc wireless networks. *Proceedings of the 2nd International Conference on Mobile Technology, Applications and Systems*, Guangzou, China, pp. 1–8.

Wengerter C, Ohlhorst J and Elbwart AV (2005) Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA. *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC)*, Stockholm, Sweden, vol. 3, pp. 1903–1907.

Witrisal K (2002) OFDM air interface design for multimedia communications. Ph.D. thesis, Delft University of Technology.

Wong CY, Cheng RS, Letaief KB and Murch RD (1999a) Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, **17**(10), 1747–1758.

Wong CY, Tsui CY, Cheng RS and Letaief KB (1999b) A real-time subcarrier allocation scheme for multiple access downlink OFDM transmission. *IEEE VTC*, Amsterdam, The Netherlands, vol. 2, pp. 1124–1128.

Wormell D, Foxlin E and Katzman P (2007) Improved 3d interactive devices for passive and active stereo virtual environments. *Proceedings of the 13th Eurographics Workshop on Virtual Environments*, Weimar, Germany.

Xiong L (1998) A selective model to suppress NLOS signals in angle-of-arrival AOA location estimation. *Proceedings of the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Boston, MA, vol. 1, pp. 461–465.

Xu G (2003) *GPS: Theory, Algorithms and Applications*, Springer.

Yoon J, Liu M and Noble B (2003) Random waypoint considered harmful. *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE INFOCOM 2003*, San Francisco, CA, vol. 2, pp. 1312– 1321.

Zander J and Erikkson H (1993) Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms. *IEEE Journal on Selected Areas in Communications*, **11**, 926–933.

Zeisberg S and Schreiber V (2008) EUWB – coexisting short range radio by advanced ultra-wideband radiotechnology. *Proceedings of the ICT Mobile Summit 2008*, Stockholm, Sweden.

Zimmermann T (1991) *Elektrooptische Entfernungsmessung mit Bandspreizverfahren*. VDI.

Ziv N and Mulloth B (2006) An exploration on mobile social networking: Dodgeball as a case in point. *Proceedings of the 5th IEEE International Conference on Mobile Business (ICMB 2006)*, Copenhagen, Denmark, p. 21.

Zuendt M, Dornbusch P, Schafer T, Jacobi P and Flade D (2005) Integration of indoor positioning into a global location platform. *First Workshop on Positioning, Navigation and Communication (WPNC 2004)*, Hanover, Germany, pp. 55–56.

Index

© 2010 John Wiley & Sons, Ltd